

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ELABORATION D'UN CADRE D'ÉVALUATION DES SYSTÈMES TUTORIELS
INTELLIGENTS: APPLICATION À ANDES2

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN INFORMATIQUE

PAR
NKOA ANTOINE NAMA TSIMI

FÉVRIER 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

À ma mère

REMERCIEMENTS

Je présente mes remerciements au Pr. Roger NKAMBOU, directeur du laboratoire GDAC (Le laboratoire de recherche sur la Gestion, Diffusion, et Acquisition de Connaissances) de l'UQAM (Université du Québec à Montréal), pour m'avoir accueilli au sein de son équipe et avoir accepté d'être mon encadreur pour la rédaction de ce mémoire. Merci de toute la confiance qu'il m'a faite, et le soutien sans faille qu'il m'a accordé.

Je remercie mes collègues Mohamed Gaha, Usef Faghihi pour leurs observations et leur soutien qui m'ont été d'une grande utilité.

Merci également à Alix Boc, Abdoulaye Boubakar Diallo du laboratoire de bio-informatique et à mon ami Maher Gaha qui m'ont aidé dans la mise en forme de ce document.

Je tiens à remercier particulièrement ma copine Marthe-Sandrine Eiymo pour tout l'appui qu'elle m'a accordé.

Mes remerciements s'adressent aussi à ma famille ainsi qu'à mes amis pour leur encouragement et leur soutien moral.

À tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet, qu'ils trouvent ici mes remerciements les plus sincères.

TABLE DE MATIÈRE

TABLE DE MATIÈRE	iv
LISTE DES FIGURES	vii
LISTE DES TABLEAUX	viii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	x
RÉSUMÉ	xi
CHAPITRE I	1
INTRODUCTION	1
1.1 Contexte	1
1.2 Problématique	3
1.3 Objectif	4
1.4 Plan	5
CHAPITRE II	7
LES SYSTEMES TUTORIELS INTELLIGENTS	7
2.1 Les composantes structurelles	7
2.1.1 Le modèle du domaine	8
2.1.2 Le modèle expert	12
2.1.3 Le modèle pédagogique	15
2.1.4 Le modèle de l'apprenant	17
2.1.5 Le modèle de communication ou interface	19
2.3 Débat sur les STI	21
CHAPITRE III	23
MÉTHODES D'ÉVALUATIONS DES SYSTÈMES TUTORIELS INTELLIGENTS	23
3.1 Les variables dépendantes d'une évaluation d'apprentissage :	24
3.2 Revue des méthodes d'évaluation	25
3.2.1 Preuve de rectitude	26
3.2.2 « Additive experimental design » (études de l'impact des composantes)	26
3.2.3 La précision du diagnostic d'un STI	26

3.2.4	Rétroaction immédiate/ instruction qualité sur l'apprenant.....	27
3.2.5	Analyse de sensibilité.....	27
3.2.6	Recherche expérimentale.....	27
3.2.7	Expert de connaissances et de comportements.....	28
3.2.8	Les expériences du magicien d'OZ.....	28
3.2.9	Évaluation basée sur critère.....	28
3.2.10	Test pilote.....	28
3.2.11	Certification.....	29
3.2.12	Évaluation extérieure (outside assesement).....	29
3.2.13	Existence de preuve.....	29
3.2.14	Observation et classification qualitative de phénomènes.....	30
3.2.15	Tâches structurées et classification quantitative de phénomènes.....	30
3.2.16	Études comparatives.....	30
3.3	Une classification des méthodes.....	31
3.4	Points de vue sur l'évaluation des STI.....	33
3.4.1	Point de vue architectural.....	33
3.4.2	Utilisabilité-utilité-acceptabilité.....	39
3.4.3	Génie logiciel.....	43
3.5	Perspectives.....	45
	CHAPITRE IV.....	47
	PROPOSITION D'UNE GRILLE D'ÉVALUATION DES SYSTÈMES TUTORIELS INTELLIGENTS.....	47
4.1	Justification et objectif de notre grille d'évaluation.....	47
4.2	Grille d'évaluation de la dimension ergonomique.....	49
4.2.1	Guidage.....	49
4.2.2	Charge de travail.....	53
4.2.3	Contrôle Explicite.....	56
4.2.4	Adaptabilité.....	59
4.2.5	Gestion des erreurs.....	61
4.2.6	Homogénéité/cohérence.....	64
4.2.7	Signifiante des codes et dénominations.....	65
4.2.8	Compatibilité.....	66

4.2.8	Compatibilité	66
4.3	Grille d'évaluation de l'utilité d'un système tutoriel intelligent.....	67
4.3.1	Le modèle apprenant.....	67
4.3.2	Le modèle pédagogique	69
4.3.3	Le modèle du domaine ou Base de connaissance	71
4.3.4	La modèle expert.....	73
	CHAPITRE V	75
	PRÉSENTATION D'ANDES.....	75
5.1	Architecture.....	75
5.1.1	Interface	76
5.1.2	Modèle pédagogique.....	78
5.1.3	Base de connaissance et modèle expert.....	85
5.1.4	Modèle de l'apprenant.....	87
5.2	Résumé de l'évaluation empirique.....	88
	CHAPITRE VI	92
	ÉVALUATION	92
6.1	Contexte de l'évaluation d'ANDES2.....	92
6.2	Résultats de l'évaluation d'ANDES2.....	92
6.2.2	Dimension utilité.....	106
6.3	Discussion de l'évaluation	111
6.3.1	Dimension ergonomique.....	112
6.3.2	Dimension de l'utilité	113
6.3.3	Analyse comparatives des résultats de notre évaluation d'ANDES2	114
	CHAPITRE VII.....	116
	CONCLUSION	116
7.1	Synthèse	116
7.2	Perspectives et limites de notre grille d'évaluation.....	117
	BIBLIOGRAPHIE	119

LISTE DES FIGURES

Figure 2.1.1 Interaction entre les composantes d'un STI (Beck, Stern et Haugsjaa, 2004)	7
Figure 2.1.2.1 les types de modèles expert (Nkambou, 2005)	12
Figure 2.1.2.1.1 Le modèle boîte noire (Nkambou, 2005)	13
Figure 2.1.2.2.1 Exemple de règle de MYCIN (Nkambou, 2005)	14
Figure 2.1.3.1 Exemple de règle socratique (Nkambou, 2005)	16
Figure 2.1.4.1.1 OverlaySimple (Nkambou, 2005)	17
Figure 2.1.4.1.2 Overlay de Type Perturbateur (Nkambou, 2005)	18
Figure 2.1.5.2.1 Interface de canadarmTutor (Nkambou, 2005)	20
Figure 3.3.1 Charte de classification des méthodes d'évaluation (Iqbal A. <i>et al.</i> , 1999)	33
Figure 3.4.1.5.1 Cadre pour l'évaluation des SHA	38
Figure 5.1.1.1 Interface ANDES2 (coupée à droite, (VanLehn <i>et al.</i> , 2005a))	76
Figure 5.1.2.2.1 Exemple de séquence de conseil d'ANDES2 (VanLehn <i>et al.</i> , 2005a)	80
Figure 5.1.2.3.1 Graphe de solution du problème de la figure 5.1.1	82
Figure 5.1.3.1 Méthode de résolution de problème pour la seconde loi de NEWTON	86
Figure 6.3.1.1 histogramme récapitulatif de l'évaluation de l'utilité d'ANDES 2	112
Figure 6.3.2.1 histogramme récapitulatif de l'évaluation de l'utilité d'ANDES 2	113

LISTE DES TABLEAUX

Tableau 3.4.2.1.1 Présentation des mesures et critères pour l'évaluation de l'utilité d'un STI (Tricot et al., 2003)	40
Tableau 3.4.2.2 Présentation des mesures et critères pour l'évaluation de l'utilisabilité d'un STI (Tricot et al., 2003)	42
Tableau 3.4.2.3 Présentation des mesures et critères pour l'évaluation de l'accessibilité d'un STI (Tricot et al., 2003)	43
Tableau 4.2.1.1.1 Évaluation de l'ergonomie : l'incitation	51
Tableau 4.2.1.2.1 Évaluation de l'ergonomie : le Groupement/Distinction par le format ou le groupement	52
Tableau 4.2.1.3.1 Évaluation du Feedback Immédiat	52
Tableau 4.2.1.4.1 Évaluation de l'ergonomie : la Lisibilité	53
Tableau 4.2.2.1.1 Évaluation de l'ergonomie : brièveté.....	55
Tableau 4.2.2.2.1 Évaluation de l'ergonomie : Densité informationnelle.....	56
Tableau 4.2.3.1.1 Évaluation de l'ergonomie : Actions explicites.....	58
Tableau 4.2.3.2.1 Évaluation de l'ergonomie : contrôle de l'utilisateur	59
Tableau 4.2.4.1.1 Évaluation de l'ergonomie : la flexibilité	60
Tableau 4.2.4.2.1 Évaluation de l'ergonomie: La protection contre les erreurs.....	61
Tableau 4.2.5.1.1 Évaluation de l'ergonomie: La protection contre les erreurs.....	62
Tableau 4.2.5.2.1 Évaluation de l'ergonomie: La qualité des messages d'erreurs.....	63
Tableau 4.2.5.3.1 Évaluation de l'ergonomie: La correction des erreurs.....	64
Tableau 4.2.6.1 Évaluation de l'ergonomie: Homogénéité et cohérence	65
Tableau 4.2.7.1 Évaluation de l'ergonomie: Signifiante des codes et dénomination	66
Tableau 4.2.8.1 Évaluation de l'ergonomie : compatibilité	67
Tableau 4.3.1 Évaluation de l'utilité : le modèle apprenant.....	69
Tableau 4.3.2.1.1 Évaluation de l'utilité : le modèle pédagogique- approche et objectifs.....	70
Tableau 4.3.2.1.1 Évaluation de l'utilité : le modèle pédagogique- adéquations des scénarios didactiques	71

Tableau 4.3.3.1 Évaluation de l'utilité : base de connaissance-conformité au domaine et facilitation de l'atteinte des objectifs pédagogiques	73
Tableau 4.3.4.1 Évaluation de l'utilité : modèle expert-rectitude	74
Tableau 5.2.1 Résultats pour l'examen d'une heure (Vanlehn et al., 2005b).	89
Tableau 5.2.2 Résultats de l'effet des tailles des examens d'une heure par critère (Vanlehn et al., 2005b).	90
Table 5.2.3 Résultats de l'effet des tailles de l'examen final (VanLehn et Al., 2005b).....	91

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CAI- Computer Assisted Instruction

CBT- Computer Base Training

NSTA- National Science Teachers Association

IHM- Interface Homme Machine

SHA- Système Hypermédia Adaptatif

STI- Système Tutoriel Intelligent

GDAC- Le laboratoire de recherche sur la Gestion, Diffusion, et Acquisition de
Connaissances

RÉSUMÉ

De nos jours, les systèmes tutoriels intelligents sont de plus en plus performants. Ils ont pour objectif de permettre à un apprenant ou groupe d'apprenants d'acquérir de nouvelles connaissances en s'adaptant à leurs besoins. Cependant, malgré leur objectif prometteur, ces systèmes ne sont pas très répandus dans le marché scolaire. De plus, comme l'a remarqué Sharon Ainsworth (Ainsworth, 2005) les STI sont « mourants ». Le problème est de savoir quels sont les critères de qualité auxquels doit répondre un STI pour pouvoir augmenter son intégration dans le milieu académique. Beaucoup d'approches d'évaluation dans la plupart des cas empiriques ont été proposées en réponse à ce problème. Toutefois, ces solutions empiriques ont pour principal défaut qu'il faut attendre une version fonctionnelle du STI. Par conséquent, elles posent un problème de temps et d'argent.

Afin de pallier ces difficultés, dans le présent document nous avons proposé une grille d'évaluation bidimensionnelle par inspection. En raison de leurs influences non négligeables sur l'acceptabilité d'un produit, les deux dimensions traitées dans notre grille seront l'utilisabilité (ou l'ergonomie) et l'utilité d'un STI. Notre grille est novatrice dans la mesure où elle est une compilation des différents critères d'évaluation. De plus, elle a été conçue pour une approche par inspection qui pourra guider la conception et permettre d'avoir une idée précise sur l'acceptation du STI avant un test empirique coûteux (en temps, en argent et en ressource). Afin de développer notre grille, nous nous sommes interrogés sur ce que l'on attend de chacune des composantes d'un STI. Par la suite, nous avons recensé les différentes techniques et approches d'évaluation STI réalisées jusqu'ici. À partir de cela, nous avons pu ressortir les critères d'utilité et d'ergonomie auxquels doit répondre un STI pour être acceptable. Enfin, pour montrer la pertinence de notre grille, nous l'avons appliquée à un STI mature ayant déjà subi cinq ans d'évaluation empirique à savoir ANDES.

Mots clés : Évaluation; STI; Utilité; Ergonomie ou utilisabilité; Qualité; Acceptabilité.

CHAPITRE I

INTRODUCTION

1.1 Contexte

De nos jours, le mode de formation par apprentissage en ligne (ou E-learning) est un mode auquel les organisations et établissements de formation recourent de plus en plus fréquemment. Ce mode de formation consiste à donner accès de façon formelle ou non formelle à des activités de formation mises en œuvre à travers un processus et des événements d'apprentissage. Il utilise tout médium électronique comme l'intranet, l'extranet, le Cd-rom, la cassette vidéo, le DVD, etc. Ce type de formation présente de nombreux avantages sur le plan formatif et socio-économique ce qui explique sa popularité actuelle (Gouadjo-Ako, 2005).

En effet, sur le plan formatif, d'une part il permet aux formateurs de communiquer plus facilement avec les apprenants (par exemple grâce à l'utilisation de plateformes virtuelles). D'autre part, il permet aux apprenants de moduler un horaire d'apprentissage conciliant la formation à leurs autres activités. Enfin, formateurs et apprenants ont désormais grâce à ce mode de formation un accès facilité à des ressources pédagogiques dispersées de qualité et diversifiées.

Sur le plan socio-économique, ce mode de formation permet aux entreprises de réduire les coûts de leurs formations et d'augmenter leur capital de savoir et de connaissances. De plus, en rendant les ressources pédagogiques disponibles à un plus grand nombre et en permettant

un accès flexible à ces ressources, cela accroît de façon générale le savoir de la société, et idéalement, contribue à sa transformation

Les systèmes de cours assistés par ordinateurs (CBT¹) et des instructions assistées par ordinateurs (CAI²) sont les premières tentatives d'utilisation de l'ordinateur comme moyen d'enseignement pour ce type de formation. Dans ces systèmes, les instructions n'étaient point personnalisées aux besoins de l'apprenant. Elles découlaient d'une approche behavioriste de l'enseignement. En d'autres termes, ce n'était que la réponse finale de l'apprenant qui importait. Ainsi, les décisions sur la façon dont l'apprenant passait à travers la matière étaient des scripts énoncés de la façon suivante : « si la question 51 est répondue correctement alors, passez à la question 75 sinon passez à la question 44 ». Les habiletés de l'apprenant n'étaient pas prises en considération par ces systèmes (Beck, Stern et Haugsjaa, 2004). De plus, ces systèmes étaient limités par une présentation rigide du matériel didactique et une absence de théorie éducative. Par conséquent, bien que CBT et CAI soient une bonne aide dans l'apprentissage, ces derniers sont très loin de fournir à un apprenant une aide équivalant à celle qu'il recevrait d'un enseignant humain. Pour pallier cela, il faudrait que ces systèmes puissent concilier ce raisonnement autant sur le domaine d'apprentissage que sur l'apprenant lui-même. Les recherches visant à résoudre ce problème ont donné naissance aux systèmes tutoriels intelligents (STI).

Les systèmes tutoriels intelligents sont des systèmes ayant pour objectif de permettre à un apprenant ou groupe d'apprenants d'acquérir de nouvelles connaissances en s'adaptant à leurs besoins. Ils s'opposent aux systèmes classiques tels les CBT ou CAI dans la mesure où ils connaissent la matière à enseigner (elle constitue leur base de connaissance), comment l'enseigner (en d'autres termes, le modèle pédagogique à utiliser) et acquérir des informations pertinentes sur l'apprenant (qui constituent son modèle de l'apprenant). Par ailleurs, un STI est capable de résoudre par lui-même les problèmes de la matière qu'il

¹ CBT : Computer Based Training

² CAI : Computer Assisted Instruction

enseigne grâce à son modèle expert. Ainsi, à travers l'interaction entre tous les modules les constituant, les STI sont capables de fournir un enseignement adapté à un apprenant sans l'intervention d'un humain.

Il y a plus d'une vingtaine voire trentaine d'années que les STI n'ont cessé d'évoluer et de se répandre. Malgré cela, ils sont loin d'envahir le cadre académique ou formatif d'entreprise. De plus, la plupart des STI existants sont des prototypes. Dans la section ci-après, nous présenterons plus en détail la problématique révélée par la situation présente des STI.

1.2 Problématique

Il est vrai que les systèmes tutoriels intelligents sont de plus en plus répandus de nos jours. En effet, les STI se retrouvent dans des secteurs d'activités variés tels : l'armée (exemple Sherlock pour apprendre à réparer les avions F16), la médecine (Guidon pour apprendre à établir de bons diagnostics), l'astronomie (CanadarmTutor pour apprendre aux astronautes la manipulation du bras canadien sur la station spatiale internationale), etc. Toutefois, malgré le fait qu'ils touchent divers secteurs d'activités de la vie courante, la plupart des STI ont encore du mal à percer le marché industriel. Cela est révélateur du fait que les STI actuels sont loin d'être parfaits d'où l'importance de leur évaluation.

En outre, de par leur objectif, les STI devraient être très répandus dans le marché scolaire, ce qui n'est pas le cas. Une des raisons de cet insuccès est liée à leur qualité. Ce fait a d'ailleurs été souligné par plusieurs chercheurs du domaine (Ainsworth, 2005 Tchounikine et al., 2002). En effet, d'une part, l'absence de cadre d'ingénierie empêche d'avoir une garantie sur la qualité des systèmes produits (Tchounikine et al., 2004). D'autre part, il n'existe pas de standard d'évaluation de ces systèmes (Ainsworth, 2005). Le problème qui se pose est de savoir quels sont les critères auxquels doit répondre un STI pour pouvoir augmenter son intégration en milieu scolaire ou en entreprise. Puisqu'il est difficile de parler réellement de méthodologie efficace sans avoir des critères de qualité clairs, nous avons choisi d'orienter nos recherches plutôt sur ce dernier point.

Par ailleurs, nous avons remarqué que dans la plupart des cas, les approches d'évaluation empiriques ont été proposées en réponse à ce problème. Ces approches empiriques ont pour principal défaut qu'au moins une version du STI a été développée. En cela, elles ne peuvent servir de guides aux développeurs ou aux concepteurs. De plus, l'approche empirique est coûteuse en matière de temps (il faut attendre qu'une version soit disponible, attendre le rapport de l'équipe d'évaluation) et d'argent (peut impliquer une réingénierie du système; la réunion d'une équipe d'évaluation a un coût monétaire et temporel). Les concepteurs et développeurs gagneraient à avoir une grille de critères leur permettant de minimiser les erreurs. De ce fait, le nombre d'évaluations empiriques serait réduit et leurs résultats seraient plus positifs.

1.3 Objectif

Dans le présent document, nous proposons une grille d'évaluation bidimensionnelle par inspection. Les deux dimensions visées par la grille seront l'utilisabilité (l'ergonomie) et l'utilité d'un STI. D'une part, l'aspect ergonomique permet à un individu de se familiariser plus facilement avec un outil ou de le rejeter. Ce facteur s'adresse au STI dans son ensemble, mais plus particulièrement à son interface de communication. L'aspect de l'utilité quant à lui concerne plus la capacité du système à permettre à l'apprenant de développer voire d'acquérir des habiletés ou connaissances. En d'autres termes, ce facteur concerne la pédagogie du système. Le côté pédagogique d'un STI est lié à l'interaction entre le modèle expert, la base de connaissance et le modèle pédagogique. L'évaluation de l'utilité consiste d'une part à s'assurer qu'une pédagogie pertinente par rapport au contexte d'utilisation est développée par le système et que d'autre part, cette pédagogie facilitera le développement des compétences de l'apprenant.

La grille d'évaluation que nous proposerons est novatrice dans la mesure où elle découlera d'une adaptation des approches et critères d'évaluation existants, mais surtout par la complétude de cette évaluation. En effet, le fait de s'interroger sur l'ergonomie et l'utilité d'un STI permet de toucher deux aspects pertinents à l'acceptabilité de ce dernier. De plus,

notre grille sera conçue pour une approche par inspection qui pourra guider la conception et permettre d'avoir une idée précise sur l'acceptation du STI avant un test empirique coûteux (en temps, en argent et en ressources).

En résumé, le travail présenté dans ce document vise à :

- 1) Proposer une grille d'évaluation couvrant les dimensions de l'utilité et de l'ergonomie d'un STI ;
- 2) Appliquer cette grille à ANDES qui est un STI mature afin d'en montrer la pertinence ;
- 3) Développer une méthodologie efficace pour le déploiement d'un STI intégrant des critères de qualité.

1.4 Plan

Pour mettre au point notre grille, notre démarche va consister dans un premier temps à voir quels sont les services attendus de chacune des composantes d'un STI. Par la suite, nous allons recenser les différentes techniques et approches d'évaluation des STI réalisées jusqu'ici. . À partir de cela, nous ressortirons les critères d'utilité et d'ergonomie auxquels doit répondre un STI pour être acceptable. Par la suite, nous allons appliquer notre grille d'évaluation à un STI mature, à savoir ANDES. L'évaluation d'ANDES qui est un STI ayant déjà subi cinq ans d'évaluation empirique nous permettra de calibrer notre grille. Enfin, nous discuterons de nos résultats et de la pertinence de notre grille d'évaluation.

En plus des chapitres d'introduction (chapitre 1) et de conclusion (chapitre 7), ce mémoire comporte cinq autres chapitres, deux annexes et une bibliographie. Plus précisément, les chapitres seront repartis comme suit :

- Dans le deuxième chapitre, nous effectuerons une présentation des STI : pour ce faire, nous présenterons leurs principales composantes et leur portée;

- Dans le troisième chapitre, nous allons analyser l'état de l'art des différentes techniques d'évaluation existantes : nous présenterons les différentes mesures pour l'évaluation d'un STI, les méthodes d'évaluation existantes, et différents points de vue sur ce qu'on doit évaluer dans un STI;
- Dans le quatrième chapitre, nous élaborerons notre grille d'évaluation : nous présenterons nos critères pour l'évaluation ergonomique d'un STI et les critères d'évaluation de l'utilité par rapport aux principales composantes d'un STI;
- Dans le cinquième chapitre nous présenterons le système de référence pour notre évaluation à savoir : ANDES. Pour ce faire, nous étudierons son architecture et ferons un bilan de ses précédentes évaluations;
- Dans le sixième chapitre, nous présenterons et discuterons du résultat de notre évaluation ;
- Dans le chapitre 7, nous ferons une synthèse de notre travail et présenterons les perspectives et limites de notre grille d'évaluation.

CHAPITRE II

LES SYSTEMES TUTORIELS INTELLIGENTS

2.1 Les composantes structurelles

Les systèmes tutoriels intelligents peuvent sembler être des systèmes monolithiques, mais pour des fins d'analyse et de conceptualisation, il est plus facile de les voir comme un ensemble de plusieurs composants interdépendants. De manière classique les composantes d'un système tutoriel intelligent sont groupées en quatre modules : le modèle de l'apprenant, le modèle pédagogique, le modèle du domaine, et le modèle de communication ou interface. Beck, Stern et Haugsjaa (2004) ont rajouté à cette architecture classique un cinquième modèle à savoir : modèle expert. Ce cinquième modèle était inclus dans le modèle de connaissance du domaine de l'architecture classique d'un STI. La figure ci-après représente l'architecture à cinq composantes des interactions entre les modules d'un STI.

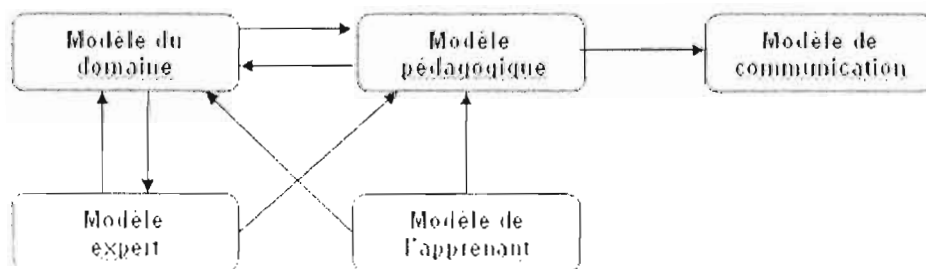


Figure 2.1.1 Interaction entre les composantes d'un STI (Beck, Stern et Haugsjaa, 2004)

Dans les sous-sections qui vont suivre nous présenterons chacune des cinq composantes d'un STI et les principales difficultés présentées par l'élaboration de chacune d'elles.

Dans les sous-sections qui vont suivre nous présenterons chacune des cinq composantes d'un STI et les principales difficultés présentées par l'élaboration de chacune d'elles.

2.1.1 Le modèle du domaine

Cette composante contient les connaissances et/ou habiletés que le STI souhaite faire acquérir à l'apprenant. C'est à partir des connaissances qu'il y a construction de tout le processus d'apprentissage, à savoir quoi enseigner et comment enseigner (Meisel et Compatangelo, 2004). En effet, le modèle du domaine appuyé par le modèle pédagogique est en mesure de présenter les connaissances adéquates à l'apprenant. Ces connaissances servent aussi à inférer³ des solutions par l'intermédiaire du modèle expert pour vérifier l'exactitude d'un exercice. Un autre modèle qui manipule les connaissances est le modèle de l'apprenant. Ce dernier enregistre des informations qui sont propres à chaque apprenant en particulier et informe sur son niveau d'assimilation.

Les connaissances ne peuvent servir que dans la mesure où l'on peut les représenter. Ainsi, dans les sous-sections suivantes nous donnerons une vision d'ensemble sur les différents types et représentations des connaissances avec leurs forces et faiblesses.

2.1.1.1 Les types de connaissances

La connaissance se définit dans l'encyclopédie psychologique de l'Agora comme *l'action ou l'acte de se faire une représentation, de s'informer ou d'être informé de l'existence de quelque chose*. La connaissance peut prendre deux dimensions complémentaires, un aspect empirique et un aspect logique. Elle est définie aussi comme un ensemble de sujets, concepts et relations touchant un domaine bien particulier. Elle a cependant historiquement été divisée en deux grandes familles, à savoir les connaissances procédurales et les connaissances déclaratives.

D'une part, les connaissances procédurales renferment des informations du type un savoir-faire. Ces connaissances sont difficilement verbalisables, et se manifestent plutôt à travers

³ L'inférence est une opération logique portant sur des propositions tenues pour vraies (les prémisses) et concluant à la vérité d'une nouvelle proposition en vertu de sa liaison avec les premières.

une action. Les connaissances procédurales sont proches de l'action et de la mise en pratique et s'acquièrent par étapes progressives. Ce type de connaissance répond à la question comment faire? (Paris, Lipson et Wixson, 1983). Par ailleurs, elles dérivent de l'emploi des systèmes experts et se basent sur l'usage des règles (p. ex. si ... alors... sinon). Ainsi, il est possible de représenter l'état des connaissances de l'apprenant par l'analyse des règles ou séquences de règles sélectionnées.

D'autre part, les connaissances déclaratives spécifient les structures de contrôle qui sont utilisées lors de la réalisation d'une action. Ces connaissances ne sont pas facilement communicables, car elles reposent sur des systèmes d'association plus ou moins complexes entre des stimulus. Dans la mémoire déclarative sont stockés des faits, des événements, des images, des concepts et des propositions accessibles sous forme de connaissances explicites. Ce genre de connaissances répond à la question quoi faire? (Paris, Lipson et Wixson, 1983).

Il est à noter que des recherches font ressortir une autre composante issue de la mémoire déclarative à savoir la mémoire épisodique. Les connaissances épisodiques font référence à des situations localisées dans un cadre spatio-temporel et s'apparentent beaucoup plus à un sous-module de la mémoire déclarative (Dubois, 2004). La mémoire épisodique utilise les mêmes zones cérébrales que celle des connaissances déclaratives (mémoire à long terme).

Maintenant que nous avons distingué les différents types de connaissances, nous nous intéresserons à la façon de les représenter.

2.1.1.2 Les principaux formalismes de représentation des connaissances

Pour retranscrire les connaissances, il existe différents formalismes possibles. Ces formalismes, selon le type de connaissances auxquels ils se réfèrent, peuvent être groupés en trois catégories à savoir : les formalismes logiques, les formalismes sémantiques et les formalismes hybrides (fusion des formalismes sémantiques et des formalismes logiques).

Les règles de production sont une représentation issue des formalismes logiques. Elles constituent l'approche la plus populaire dans le domaine de la représentation des connaissances et a été utilisée dans les systèmes experts tels MYCIN⁴. Ce type de représentation est adapté pour les connaissances heuristiques déduites de l'expérience. Les règles de production sont de la forme: **SI** condition **ALORS** conclusion. Dans cette approche, les conditions représentent des faits ou des prédicats ayant des valeurs vraies ou fausses. L'avantage de cette approche est le fait que les règles sont modulables, simples et uniformes. Ainsi, l'ajout et la suppression de nouvelles connaissances sont faciles et sans répercussion sur l'ensemble de l'application. Toutefois, cette représentation est limitée par le fait que la base de règles n'est pas structurée. En outre, il n'est pas possible d'exprimer toutes les connaissances de la forme si condition alors conclusion. En effet, les règles expriment une connaissance apparente et risquent de cacher le raisonnement profond de l'humain.

Pour remédier aux limites sus-mentionnées, de nouvelles variantes de la logique non conventionnelle (par opposition à la logique classique) ont été élaborées. Le but de ces nouvelles représentations était de tenir compte des faits ignorés par la logique classique. Ainsi, on peut citer la logique floue qui est une variante de la logique classique. Elle tire son essence de la théorie des ensembles. Elle a été développée au début des années 1960 et introduit le concept d'incertitude dans les informations. Elle offre la possibilité d'évaluer un paramètre en disant simplement dans quelle mesure il est vrai⁵. Une telle approche peut être combinée avec d'autres méthodes de représentation des connaissances et permet une meilleure gestion de l'imprécision. Des recherches ont été entreprises afin de réaliser des systèmes tutoriels intelligents à base de logique floue (Nedic, Nedic et Machotka, 2002 Nkambou, 1997, 1998).

Mise à part la logique non conventionnelle, d'autres approches ont été explorées afin de permettre une représentation de connaissances plus complexes de par leur structuration. Les

⁴ Système expert utilisé par le STI GUIDON permettant de diagnostiquer les maladies : voir Clancey, W.J. (1987). Knowledge-Based Tutoring : The GUIDON Program. Cambridge Mass: MIT Press

⁵ Contrairement à la logique booléenne, la logique floue permet à une condition d'être représentée par des valeurs autres que VRAI ou FAUX, mais à travers des degrés de vérifications compris entre [0,1].

réseaux sémantiques utilisés dans les travaux de compréhension de la langue naturelle (Quillan, 1968), constitue une approche intéressante pour représenter la connaissance. Il s'appuie sur un concept graphique pour abstraire l'information pertinente et se concentrer plutôt sur la topologie du problème. Le graphe constitue un espace de problème dans lequel on « offre la possibilité de délimiter des sous-ensembles de noeuds et d'arcs appelés des espaces » (Frebourg, 2001). SCHOLAR (Carbonell, 1970) est un exemple classique de STI se servant de réseaux sémantiques (Nkambou, 2005). Les réseaux sémantiques sont plus souples que la représentation à travers les règles de productions cependant ses propriétés formelles sont moins bien assises et moins bien étudiées.

Par ailleurs, la représentation hybride est issue de la convergence de deux types de représentations, à savoir la représentation formelle adaptée à la résolution de problème et à la représentation sémantique. Le réseau bayésien est une des représentations les plus connues issues de ce formalisme. Il est avant tout un modèle graphique dans lequel les connaissances sont représentées sous forme de variables appelées noeuds. Tous les noeuds du graphe sont reliés par un lien de dépendance direct (appelé causalité). L'absence ou la présence d'un lien informe sur l'existence ou la non-existence d'une dépendance directe. Pour chaque lien, il existe un poids qui représente la probabilité conditionnelle des noeuds. Les réseaux bayésiens constituent une solution très intéressante dans la modélisation de l'apprenant (Dubois, 2004). Ils permettent de prendre en compte des données incomplètes, incertaines et imprécises. Il est utilisé dans PrologTutor qui est un STI servant à l'apprentissage du langage Prolog (Tchetagni, Nkambou et Bourdeau, 2005).

En définitive, nous pouvons dire que l'importance du modèle d'un domaine vient du fait qu'il gère les connaissances du système. Les systèmes tutoriels intelligents sont fortement tributaires de la connaissance qui les constitue. Toutefois, c'est de la représentation de la connaissance (formalisme utilisé, niveau de granularité, expressivité des connaissances) que dépendra en grande partie l'efficacité de ces systèmes.

2.1.2 Le modèle expert

Ce modèle permet au système d'acquérir des compétences sur la matière enseignée. Ainsi, grâce à ce modèle, le système peut : résoudre les problèmes soumis à l'apprenant, suivre le cheminement du raisonnement de l'apprenant et s'adapter aux différentes singularités de raisonnement de chaque apprenant. En cela, le modèle expert des STI apporte la singularité évolutive qui le différencie des anciens CBT dans la mesure où ici l'apprenant n'a pas un cheminement prédéfini dans le système.

Cette aptitude à s'adapter et à suivre le raisonnement de l'apprenant a fait du modèle expert une composante essentielle de tout STI. Cela représente 50% des efforts de développement. En outre, pour pouvoir réaliser cette aptitude, ce modèle doit implémenter un modèle permettant de raisonner sur les connaissances de l'apprenant. A ce jour, plusieurs modèles de représentation sont utilisés : trouver une méthode de raisonnement sur le domaine (approche « boîte noire »); développer un système expert; simuler la façon dont l'humain utilise les connaissances (modèle cognitif). Comme le montre la figure ci-après, l'efficacité pédagogique et l'effort d'implantation augmentent à mesure que l'on se rapproche du raisonnement humain.

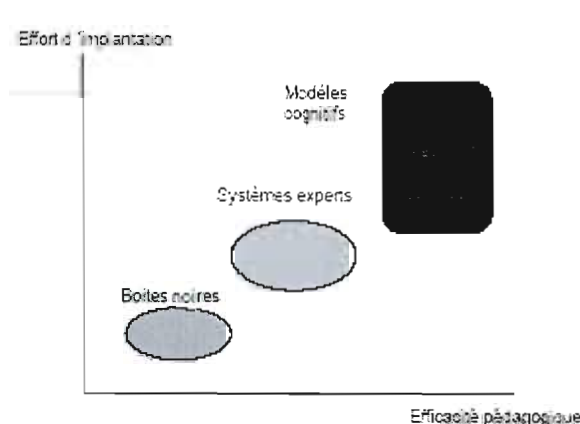


Figure 2.1.2.1 les types de modèles expert (Nkambou, 2005)

2.1.2.1 Le modèle de boîtes noires



Figure 2.1.2.1.1 Le modèle boîte noire (Nkambou, 2005)

Comme nous le montre la figure ci-dessus, cette approche propose un raisonnement sur les entrées et les sorties, mais pas les explications du problème. En conséquence, les tuteurs résultants sont réactifs (vrai, faux, faire ceci). Le système SOPHIE (Brown, Burton et Bell, 1975) est un exemple; il utilise le simulateur SPICE de circuits électroniques à l'aide d'un modèle mathématique. Il travaille sur un ensemble d'équations, signale les erreurs de l'apprenant, mais ne peut pas expliquer ses décisions. Il est pourtant important de pouvoir interrompre l'apprenant pour lui expliquer le problème.

En conclusion, pour pouvoir parvenir à fournir à l'apprenant des solutions plus raffinées, il faut pouvoir avoir accès au raisonnement sur les connaissances internes du système, ce qui permettra d'expliquer les erreurs.

2.1.2.2 Le modèle en boîte de verre

Il consiste en l'utilisation de systèmes experts. Un système expert est construit par des cognitivistes et experts du domaine. Il se base sur la représentation des connaissances pour formaliser et raffiner les concepts. Grâce au système expert, le STI peut en comparant le

raisonnement⁶ de l'apprenant à celui de l'expert, indiquer avec précision où l'apprenant a eu des difficultés.

GUIDON (Clancey, 1987) est un exemple classique de STI utilisant un système expert. En effet, il est basé sur Mycin (système permettant de diagnostiquer des maladies infectieuses) lequel utilise des règles (450 au total) définies sur les différences entre l'expert et l'apprenant. La figure ci-après présente un exemple de règle tel que défini dans GUIDON.

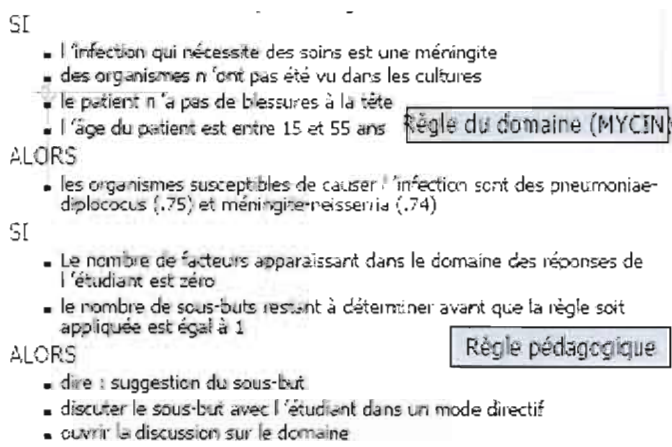


Figure 2.1.2.2.1 Exemple de règle de MYCIN (Nkambou, 2005)

2.1.2.3 Le modèle cognitif

Le but de ce modèle est de simuler la résolution d'un problème dans un domaine où la connaissance est décomposée de manière communicable par le STI. Il s'appuie sur le modèle de tutorat humain et les théories de la cognition.

Ce modèle est onéreux en termes de temps, de ressources (ordinateurs performants) et de techniques avancées (par exemple l'implantation de l'utilisation du langage naturel) qu'il

⁶ Ici le raisonnement se réfère au processus de faire coopérer des connaissances, des stratégies de résolution de problèmes avec l'objectif d'atteindre une conclusion.

nécessite. Toutefois, les progrès en sciences cognitives, la performance croissante des ordinateurs, permettent de relativiser ces coûts.

Les STI basés sur la théorie ACT sont des exemples de tuteurs cognitifs. En effet, selon la théorie de cognition humaine (ACT-R) de Anderson : "L'acquisition des habiletés cognitive se réalise par les règles de productions". Geometry Tutor et LISP Tutor découlent de cette théorie. Ils se comportent comme des guides de résolution de problème. Pour ce faire, ils utilisent la technique de traçage (analyse étape par étape) du raisonnement de l'apprenant à partir de règles de production réparties en deux sous-ensembles (correctes ou erronées). Ils préconisent la méthode d'intervention immédiate en cas de problème, car selon les auteurs, la correction différée est beaucoup plus coûteuse en temps (Nkambou, 2005)

2.1.3 Le modèle pédagogique

Le modèle pédagogique sert à instrumenter l'intervention pédagogique du système. Il utilise les informations du modèle de l'apprenant pour planifier et déterminer les aspects du modèle de connaissances (ou modèle du domaine) à présenter à l'apprenant de façon adéquate. En conséquence, ce modèle doit être capable de: modéliser l'interaction entre le tuteur et l'apprenant; concevoir les tâches exposées à l'apprenant de façon adaptative; fournir des rétroactions et de l'aide adaptée à l'apprenant (p. ex., il pourrait fournir de la rétroaction de façon proactive ou réactive); de mener des activités pédagogiques telles qu'accorder des explications ou des conseils et/ou faire passer des tests.

L'objectif final visé par ce modèle est de minimaliser les différences entre les connaissances de l'expert du domaine et l'apprenant. Pour ce faire, les interventions potentielles du système s'appuient sur des stratégies pédagogiques dont les principales sont les suivantes :

- Le « coaching » : il consiste à conseiller et guider l'apprenant lorsqu'il s'éloigne de la solution (Nkambou, 2005). Ce mécanisme détermine le moment convenable pour interrompre l'apprenant et intervenir. Cette intervention peut avoir des formes aussi

variées que : un avertissement pour signaler à l'apprenant qu'il a pris une mauvaise décision ou un encouragement lorsqu'il répond correctement à une question, un conseil pour remédier à une situation précise, des renseignements détaillés, etc....

- L'apprentissage par perturbation : il peut se faire en employant un agent perturbateur qui va parfois fournir des réponses correctes et d'autres fois, essayer d'induire l'apprenant en erreur en lui donnant de mauvaises réponses. Le but ici est de vérifier et d'améliorer la confiance et l'estime de soi chez l'apprenant.
- L'apprentissage socratique : il consiste à questionner l'apprenant sur ses erreurs pour le faire progresser. La figure ci-après nous donne un exemple de règle socratique.

SI
L'étudiant donne une explication d'un ou plusieurs facteurs qui est insuffisante
ALORS
Formuler une règle générale pour démontrer que les facteurs sont suffisants.
Demander à l'étudiant si la règle est vraie.

Figure 2.1.3.1 Exemple de règle socratique (Nkambou, 2005)

- Le guidage : cette stratégie correspond aux environnements de type micromondes. Ce type d'environnement sera détaillé dans la sous-section modèle de communication.

Il faut noter que l'implantation des stratégies pédagogiques pose des difficultés non négligeables. Par ailleurs, des recherches sont en cours pour que les STI puissent utiliser plusieurs stratégies pédagogiques. Toutefois, ces recherches sont confrontées au problème du fort lien qui existe entre la stratégie pédagogique et la représentation des connaissances incluse dans le modèle du domaine et utilisée par le modèle de l'apprenant.

2.1.4 Le modèle de l'apprenant

Le modèle de l'apprenant stocke les informations relatives à chaque apprenant individuellement. Il constitue ainsi une simulation cognitive de l'apprenant. De ce fait, ce modèle doit permettre un diagnostic et une évaluation de la compréhension du domaine par l'apprenant, de son degré de connaissance et des processus de raisonnement qu'il applique pour une adaptation dynamique et individualisée des interventions du système (p. ex. élaborer des rétroactions adéquates pour chaque type d'erreurs). Conséquemment, ce modèle est consulté régulièrement par le modèle du domaine et le modèle pédagogique pour adapter la formation aux besoins de l'apprenant. De plus, ce modèle de l'apprenant permet d'identifier des connaissances naissantes chez l'apprenant.

Par ailleurs, pour Nkambou (Nkambou, 2005), ce modèle apprenant représente l'état des connaissances de l'apprenant et la recherche de cet état s'apparente à un processus de diagnostic. Aussi, pour la réalisation de ce modèle plusieurs approches sont nécessaires :

2.1.4.1 Le modèle cognitif

Il est encore appelé modèle de recouvrement ou « overlay ». Il existe trois principales variantes de ce type de modèle:

- i. *Overlay de type simple* : Il se caractérise par le fait que les connaissances de l'apprentissage sont des sous-connaissances de l'expert. Ainsi, les performances de l'apprenant sont comparées à celles de l'expert dans les conditions semblables (voir figure ci-après).



Figure 2.1.4.1.1 OverlaySimple (Nkambou, 2005)

Afin de réaliser ce modèle, plusieurs approches d'implémentation existent. On peut citer parmi ces dernières : les réseaux sémantiques, l'ensemble de compétence acquise par l'apprenant, les réseaux bayésiens (p. ex. PrologTutor, Tchétagni, Nkambou et Bourdeau, 2005).

- ii. *Overlay de type différentiel* : Ce modèle est une extension du modèle précédent. Il se particularise par le fait qu'il partitionne les connaissances du domaine en connaissances déjà connues et celles qui n'ont pas encore été présentées à l'apprenant.
- iii. *Overlay de type perturbateur* : encore appelé modèle de type « buggy », ce modèle représente les erreurs d'apprentissage constatées lors de l'interaction avec l'apprenant (Voir figure ci-après).

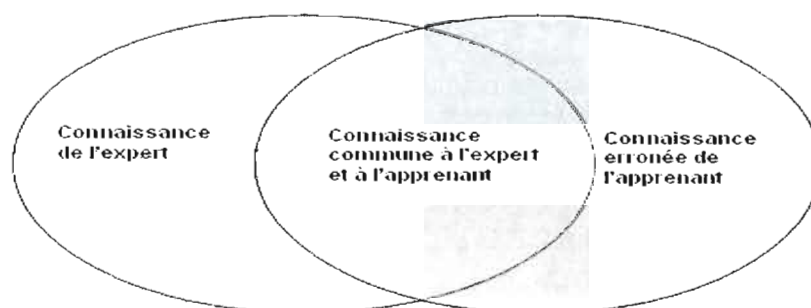


Figure 2.1.4.1.2 Overlay de Type Perturbateur (Nkambou, 2005)

Il a pour objectif de comprendre l'origine de l'erreur et de proposer une stratégie pour remédier à la solution. Pour ce faire, il inclut des connaissances ne figurant pas dans le modèle du domaine du tuteur telles qu'une description des erreurs de mauvaise compréhension en relation avec l'expert du domaine.

2.1.4.2 Le modèle affectif

Il permet le stockage des préférences de l'apprenant, de son profil psychologique et émotionnel, etc. En d'autres termes, il s'agit de tenir compte des informations pouvant influencer le comportement des apprenants en plus des connaissances du domaine des

connaissances de ce dernier. Les informations ainsi récoltées sur l'apprenant peuvent évoluer dynamiquement (p. ex. dépendamment de l'état émotionnel) et en fonction du contexte (p. ex. la motivation). Comme le souligne Nkambou (Nkambou, 2005) ces informations permettent au tuteur d'adapter son interaction en fonction de l'apprenant.

2.2.4.3 Le modèle inférentiel

Ce modèle s'intéresse particulièrement au diagnostic par l'inférence en relation avec la cause des erreurs de l'apprenant tels que : le « modèle tracing » (qui consiste à suivre et analyser la trace des activités de l'apprenant) et le « knowledge tracing » (qui consiste à analyser un événement d'apprentissage dans le but de déterminer les connaissances qui y ont été employées).

Lors de la conception d'un système tutoriel intelligent, il est important de tenir compte du fait que c'est l'apprenant qui joue le rôle le plus essentiel. Aussi, tous ces modèles utilisés sont intéressants dans la mesure où ils permettent de se faire une idée plus précise de l'apprenant. L'idéal serait d'avoir un modèle qui inclut les points forts des différents modèles. Toutefois, comme le rappelle Retschitzki, sans la participation active, la motivation et l'envie d'apprendre de l'apprenant, aucun logiciel ne pourra lui transmettre des connaissances de façon significative (Retschitzki, 1995). En cela, il est important d'avoir une interface de communication stimulante pour l'apprenant.

2.1.5 Le modèle de communication ou interface

L'interface sert de couche de communication entre l'apprenant et le STI. Son élaboration est basée sur des principes d'ergonomie de l'interaction afin de simplifier la communication entre le système et l'utilisateur. De nos jours beaucoup d'efforts sont mis dans le raffinement des interfaces en fonction de l'environnement d'apprentissage. Ainsi, il existe plusieurs types d'interfaces parmi lesquels nous pouvons citer : le micromonde, la simulation et l'hypermédia.

2.1.5.1 Le micromonde

Les micromondes permettent à l'apprenant de développer et simuler ses propres modèles du monde réel. À cet effet, Seymour Papert fortement inspiré des travaux de Piaget, suggère qu'il est plus important de permettre aux enfants d'apprendre à développer et à corriger leurs propres théories que de leur enseigner des théories considérées correctes. Dans LOGO, les apprenants programment une « tortue » pour construire et explorer des micromondes (Papert, 2003).

2.1.5.2 La simulation

La simulation s'impose dans des domaines à forte composante évolutive, surtout lorsque cette évolution peut-être perturbée par des événements extérieurs. Dans les Simulations, l'accent est davantage mis sur la simulation d'un modèle que sur sa construction par l'apprenant. L'apprenant apprend en modifiant des paramètres et en observant les conséquences de ses actions dans l'environnement simulé. Le CanadarmTutor est un STI permettant aux astronautes l'apprentissage de la manipulation du bras canadien dans l'espace. Ce logiciel utilise la vision des 13 caméras dont disposent les astronautes pour déplacer le bras canadien dans l'espace. L'astronaute doit être capable de choisir les meilleures vues et de déplacer le bras en évitant toute collision comme on peut le voir dans la figure ci-après. (Nkambou et al., 2005).

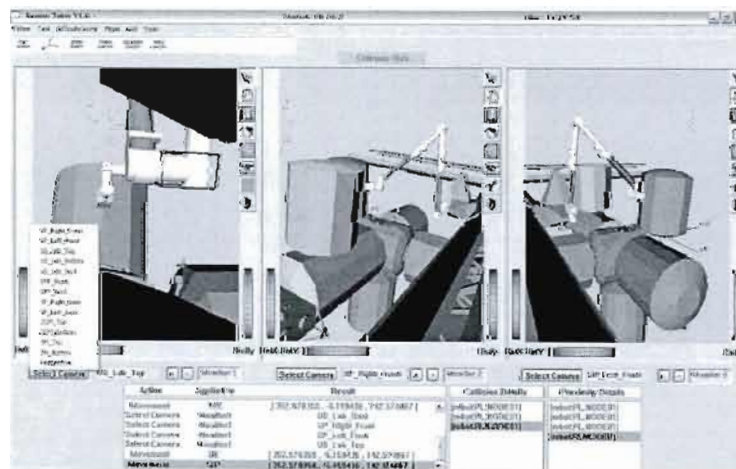


Figure 2.1.5.2.1 Interface de canadarmTutor (Nkambou, 2005)

Les problèmes des interfaces d'apprentissage sont des problèmes d'interaction personne-machine. La qualité de l'interaction peut influencer le résultat de l'apprentissage. Ainsi, les facteurs importants pour l'interface d'un STI sont :

- Utilisabilité : il faudrait que la charge mentale liée à la manipulation de l'interface soit négligeable;

- Utilité: une interface doit permettre un accès facile aux éléments primitifs du domaine d'apprentissage et supporter la métacognition de l'apprenant.

Les tendances actuelles dans l'élaboration des interfaces prônent l'utilisation du langage naturel et l'intégration de la dimension affective dans l'interaction. L'utilisation du langage naturel permettrait à l'apprenant d'utiliser ses propres termes, ce qui traduirait son niveau de compréhension de la matière. Toutefois, les tests effectués jusqu'à présent ne sont pas encore convaincants sur le gain fourni par l'utilisation du langage naturel qui est très complexe à réaliser (Aleven, Popescu et Koedinger, 2001). L'intégration de la dimension affective dans les STI s'est traduite par l'ajout de multimédia tels : des sons, des vidéos ou de personnages pour s'émouvoir. Cette intégration a pour but d'encourager l'apprenant dans ses bonnes entreprises et de le coacher dans ses erreurs comme le ferait un enseignant humain. Le principal défi dans l'ajout de cette dimension émotive est de ne pas tomber dans l'excès qui ferait de lui un facteur irritant.

2.3 Débat sur les STI

Les techniques de programmation et les ordinateurs n'ont cessé de s'améliorer ces dernières années. De plus, l'usage des ordinateurs est de plus en plus répandu de nos jours. Ces progrès ont permis une prolifération et une amélioration des STI. Cependant, ils ont également donné naissance à plusieurs débats concernant leur utilité et leur efficacité :

- « *Le degré de contrôle de l'apprenant : Jusqu'à quel point l'apprenant doit-il avoir le contrôle?* »

- *Apprentissage collaboratif vs individuel : Les apprenants doivent-ils interagir de façon individuelle ou collective ?* » (Gouadjo-Ako, 2005).
- **Réalité virtuelle et apprentissage:** L'impact de la réalité virtuelle sur l'apprentissage est-il pertinent?

Ces débats sus-mentionnés sont révélateurs du fait qu'il n'y a pas d'ingénierie établie sur la conception d'un STI. En raison de cela naît le problème de savoir: que faut-il évaluer pour juger de la bonne qualité d'un STI? Ce problème est de plus accentué par le fait que les STI tardent à percer le marché académique et industriel comme le souligne Ainsworth (Ainsworth, 2005). Dans le chapitre quatre nous apporterons une solution à cette question après avoir fait l'état de l'art sur l'évaluation des STI dans le chapitre qui suit.

CHAPITRE III

MÉTHODES D'ÉVALUATIONS DES SYSTÈMES TUTORIELS INTELLIGENTS

Malgré le fait qu'ils touchent divers secteurs d'activités de la vie courante, la plupart des STI sont encore à leur version prototype. Cela est révélateur du fait que les STI actuels sont loin d'être parfaits d'où l'importance de leur évaluation.

En effet, l'évaluation est intéressante sur un triple aspect : théorique, éducatif, et économique. Sur le plan théorique, l'évaluation des STI permet de relever des erreurs conceptuelles, d'augmenter la durée de vie des logiciels et leur utilité comme le signale Iqbal (Iqbal A. et al., 1999). Sur le plan éducatif, l'évaluation des STI permettrait d'améliorer les résultats et l'efficacité de l'apprentissage. Par ailleurs, un STI bien évalué est rentable sur le plan économique car il permettrait de rejoindre les besoins des apprenants. Ce qui aurait pour effet d'augmenter la valeur marchande de ces systèmes. De plus, les apprenants seraient plus enclins à utiliser un STI qui a passé et réussi une évaluation, car c'est une garantie de qualité. L'évolution des STI passe de ce fait par leur évaluation.

Cependant, s'il est simple de trouver des justifications à l'évaluation des STI, l'évaluation en elle-même est une tâche très complexe. En effet, le fait que ces systèmes fusionnent en leur sein plusieurs disciplines à savoir : les sciences de l'éducation, la didactique, la psychologie, les sciences cognitives, l'informatique, et l'ergonomie, contribue à complexifier cette évaluation. En raison de cela, il existe plusieurs méthodes, contextes et paramètres dont il faut tenir compte pour concevoir l'évaluation des STI.

Dans la suite du document, nous présentons un état de l'art sur l'évaluation des STI. En premier lieu, nous passerons en revue différentes variables communes pouvant servir à l'évaluation des STI. Par la suite, nous présenterons une classification de ces méthodes suivie de différents points de vue sur l'évaluation d'un STI. Enfin nous concluerons avec les perspectives de l'évaluation des STI.

3.1 Les variables dépendantes d'une évaluation d'apprentissage :

Lorsqu'on évalue un STI, le critère d'apprentissage est un des plus importants qu'on ne peut négliger. Dans cette section nous décrirons plusieurs variables (ou mesures) dont on doit tenir compte lors de l'évaluation de l'apprentissage avec ces systèmes. Nous nous baserons sur la typologie de Shaaron Ainsworth (Ainsworth, 2005) pour recenser ces variables.

- Une variable intéressante lors de l'évaluation de l'apprentissage est le gain d'apprentissage réalisé avec l'usage du STI. Pour la mesurer, il suffit par exemple d'évaluer le niveau de l'apprenant avant et après l'utilisation du STI.
- L'efficience de l'apprentissage est une mesure qui permet de déterminer si le STI est capable de réduire le temps d'apprentissage. Ce critère a été très pertinent pour l'évaluation de Sherlock (Lesgold *et al.*, 1992). En effet, Sherlock avait été intéressant pour l'armée de l'air américaine, car 20 à 25 heures passées à l'utiliser équivalaient à 4 ans d'expérience (selon les auteurs).
- Une autre mesure pertinente pour l'apprentissage est de savoir comment le système est utilisé en pratique et le profil des individus qui y ont recours. Elle permet entre autres de classifier par catégorie les utilisateurs du STI et leurs préférences, ceci permet d'éviter les résultats erronés reliés à l'apprenant. À cet effet Aleven *et al.* (2001) ont rapporté que dépendamment de l'utilisateur le résultat peut être biaisé. C'est ainsi que ces auteurs ont reconnu que leurs résultats ne reflétaient pas la réalité, car les utilisateurs du test étaient d'un niveau bien supérieur aux utilisateurs finaux du tuteur (Lesgold *et al.*, 1992).

- L'attitude de l'apprenant doit être prise en compte lors de l'évaluation de l'apprentissage. En effet, la motivation et l'état affectif peuvent influencer le résultat de son apprentissage.
- L'aptitude de l'apprenant à pouvoir expliquer, voir enseigner ce qu'il a appris permet de déterminer son niveau d'apprentissage. En effet comme le dit le dicton « ce qui se conçoit bien s'énonce clairement et les mots pour le dire arrivent aisément ».
- Les données d'interaction sont des mesures facilement récupérables et non négligeables lors de l'évaluation des STI. Ces données qui constituent la trace de l'utilisation du système par l'apprenant, aident à voir les performances de ce dernier pendant son usage du système. Parmi les données d'interaction de l'apprenant on a : le temps consacré à la réalisation d'une tâche, la progression à travers le curriculum du système, l'utilisation des outils du système (ex : l'utilisation du glossaire, de l'éditeur, ...), la performance de l'utilisateur (nombre de bonnes et mauvaises réponses), le nombre de fois qu'il a utilisé le bouton d'aide. Baker et al. (2003) ont comptabilisé le temps passé par les apprenants à résoudre les problèmes proposés par leur STI. À partir de cela, ils ont déterminé les aptitudes les plus difficiles à acquérir pour la réalisation du graphe de nuages de points.

La liste des variables dépendantes sus-citées n'est pas exhaustive. D'autres mesures peuvent être utilisées pour l'évaluation, ceci du moment qu'elles soient intéressantes. Cependant, il faut noter que pour effectuer une évaluation pertinente l'utilisation de plusieurs variables est nécessaire (Ainsworth, 2005).

3.2 Revue des méthodes d'évaluation

Les méthodes d'évaluation pour les STI sont nombreuses et doivent tenir compte de plusieurs facteurs. Ainsi, l'évaluation n'est pas une tâche triviale; comme Mark et Greer le mentionnent : « Il y a peu d'accord sur les standards dans la communauté des STI pour guider les chercheurs qui souhaitent évaluer les systèmes. Cependant, d'autres secteurs ont

développé des méthodes d'évaluation qui peuvent être applicables aux STI». (Mark et Greer, 1993). Les méthodes discutées dans cette section sont suivies par les chercheurs des différentes disciplines impliquées dans l'élaboration des STI. Les descriptions de chacune de ces méthodes ont été tirées de l'article « A classification of evaluation methods for intelligent tutoring system» (Iqbal A. *et al.*, 1999).

3.2.1 Preuve de rectitude

Elle consiste à vérifier si le système remplit les objectifs désirés. En d'autres termes, cette méthode vérifie s'il y a une correspondance entre les spécifications, la structure et le comportement du système. Les programmes conventionnels des ordinateurs sont souvent vérifiés à travers des preuves formelles. Cependant comme le mentionne Partridge cette technique est limitée par le fait que dans les STI on fait face à des problèmes analytiquement insolubles représentés comme des fonctions partiellement spécifiées (Partridge, 1986).

3.2.2 « Additive experimental design » (études de l'impact des composantes)

Des comparaisons ont été proposées pour évaluer l'impact des grandes composantes des STI qui peuvent être expérimentalement modifiées ou interrompues (O'Neil et Baker, 1987). Ces propositions ont été utilisées pour l'évaluation des composantes d'un STI par Sleeman pour étudier les différentes approches pédagogiques dans Pixie Tutor (Sleeman *et al.*, 1989). Cette approche requiert un grand nombre d'apprenants et n'est pas rentable si la composante STI est mineure. L'avantage de cette approche est d'inclure la particularité d'un STI afin d'évaluer son importance (Iqbal A. *et al.*, 1999)

3.2.3 La précision du diagnostic d'un STI

Elle peut être estimée par des procédures désignant la qualité des micro-théories utilisées par le système. Certains auteurs contestent le fait que cette méthode ait plus de prééminence sur une évaluation interne d'un STI, car les interactions pédagogiques subséquentes sont dépendantes de la bonne reconnaissance et de l'interprétation de l'erreur de l'apprenant (Legree, Gillis et Orey, 1993). Cette méthode permet un examen contrôlé du modèle à travers des conditions pré-établies et appartient donc à la recherche expérimentale (Iqbal A. *et al.*, 1999).

3.2.4 Rétroaction immédiate/ instruction qualité sur l'apprenant

La rétroaction peut être estimée à travers des procédures d'analyses séquentielles qui sont intéressantes pour une évaluation externe. Cheung et al. ont utilisé cette technique pour déterminer l'efficacité de l'apprentissage avec Smart Tutor (Cheung et al., 2003). Une façon de mesurer ce paramètre est de calculer la probabilité séquentielle, la probabilité séquentielle étant le ratio des actions d'une catégorie spécifique sur le nombre total d'actions apparaissant dans une fenêtre donnée suivant une action ciblée. L'avantage de telles mesures est que le calcul peut être effectué à différents niveaux de détail de la rétroaction d'un STI (Iqbal A. et al., 1999).

3.2.5 Analyse de sensibilité

Le but de cette méthode est de voir la diversité des réponses obtenues de la composante ou du système en fonction des informations fournies. En d'autres termes, avec cette méthode une composante ou un système peut être examiné pour voir la sensibilité de son comportement face aux informations qui lui sont fournies. Cette méthode est intéressante pour les STI dans la mesure où avoir des instructions individualisées est un objectif primordial pour les tuteurs intelligents (Mark et Greer, 1993).

3.2.6 Recherche expérimentale

Elle permet aux chercheurs d'obtenir la relation entre un ensemble d'interventions et le résultat. De ce fait, cette technique est tout à fait appropriée pour mesurer les effets de l'enseignement. Cette méthode a servi à l'évaluation de Sherlock (Lesgold et al., 1992). Par ailleurs, Mark et Greer (1993) ont défini plusieurs types de conceptions expérimentales pour mener à bien une recherche expérimentale. Elles sont classées comme suit: conception d'un groupe unique (single group design), le groupe de contrôle, et la conception quasi expérimentale. La recherche expérimentale est plus appropriée pour une évaluation externe, car elle s'attarde plus aux conclusions générales qu'à l'acquisition d'information (Iqbal A. et al., 1999).

3.2.7 Expert de connaissances et de comportements

Cette approche consiste à déterminer si un système rencontre un niveau explicite standard de performance. Elle est souvent utilisée lors de l'évaluation de la base de connaissance d'un système. L'exemple le plus connu de cette approche est le « *Turing Test* » qui compare le comportement de la machine à celui de l'humain (Parry et Hofmeister, 1986). De telles inspections sont utiles à l'évaluation spécifique de composantes de STI. Ainsi, cette approche est applicable pour une évaluation interne. Cette approche est dite exploratrice, car on n'a pas d'hypothèses de contrôle en place (Iqbal A. et al., 1999).

3.2.8 Les expériences du magicien d'OZ

Elles utilisent un humain pour simuler le comportement du système proposé. De cette façon, les aspects conceptuels du programme peuvent être testés avant leur implémentation. Cette méthode est limitée par l'imprévisibilité des réactions humaines. Par conséquent, il est difficile de déterminer comment les paramètres peuvent être observés. Cette méthode fait partie de la recherche exploratrice (Iqbal A. et al., 1999).

3.2.9 Évaluation basée sur critère

Cette évaluation consiste à effectuer des vérifications à l'aide d'une liste de critères prédéfinis. Cette méthode peut être utilisée autant pour évaluer les composantes du système que le système dans son intégralité et est par conséquent applicable autant pour une évaluation interne qu'externe (Iqbal A. et al., 1999).

3.2.10 Test pilote

Il est utilisé pendant la conception du système afin d'anticiper les problèmes dans le système. Trois types de test pilote ont été identifiés par Gagné comme fournissant des opportunités d'apprendre sur différentes parties du processus du développement. Cet apprentissage peut se faire tant au niveau des composantes qu'au niveau du système dans son ensemble (Gagne, Briggs et Wager, 1988). Les tests pilotes individuels sont utilisés au début du développement

du système. Plus tard dans le développement, les petits groupes de tests sont utilisés avec un groupe représentatif d'apprenants une fois que le système a commencé à se stabiliser. Enfin, les champs de test (Field testing) prennent place vers la fin du développement quand le système est presque terminé (Iqbal A. et al., 1999).

3.2.11 Certification

Elle est basée sur les méthodes appliquées pour définir des enseignants humains compétents. De telles procédures bénéficient ainsi de l'endossement d'une autorité de rectitude du STI. En raison de cela, elle entre dans le cadre de la recherche prospective. L'applicabilité de cette approche à l'ensemble du système la rend adéquate à une évaluation externe. Mark et Greer (1993) privilégient cette méthode comme étant d'une part un moyen d'obtenir un feed-back sur les forces et faiblesses d'un STI dans l'évaluation formatrice et d'autre part une façon d'obtenir un taux adéquat dans l'évaluation quantitative (Iqbal A. et al., 1999).

3.2.12 Évaluation extérieure (outside assesement)

L'évaluation des experts, ou d'un grand nombre d'apprenants (potentiels) du système, connue comme évaluation extérieure peut être significative dans la mesure où il existe un accord commun entre les différentes parties. D'après Murray il existe deux façons de faire cette évaluation : « *on-site expert* » et « *panels of experts* » (Murray, 1993). La première est la méthode normale qui consiste à faire observer le comportement du système par les experts. Par exemple, plusieurs enseignants font une estimation de l'efficacité d'un STI avec les apprenants qui l'utilisent. La seconde méthode s'appuie sur un groupe d'experts sélectionnés et questionnés par correspondance ou dans une assemblée. En raison du fait que cette approche opère sur le STI dans son ensemble et est basée sur l'exploration des opinions des apprenants, elle est applicable pour des évaluations prospectives externes.

3.2.13 Existence de preuve

Les conclusions de cette approche se basent sur le succès de l'implantation d'un système ou d'une méthode pour proposer une nouvelle architecture du système. Du fait que cette méthode s'applique à l'ensemble du système, elle entre dans la catégorie d'évaluation

externe. Cette méthode est efficace seulement si le chercheur décrit les objectifs et hypothèses d'étude, discute des changements conceptuels et documente les surprises et les échecs. Elle est utile dans la recherche prospective, car elle force à définir les objectifs-clés et peut servir de base pour de nouvelles conceptions.

3.2.14 Observation et classification qualitative de phénomènes

Le but de cette méthode est d'identifier les classes de phénomènes, les modèles et tendances des apprenants avec les STI. Ces données sont recueillies pour la recherche prospective principalement à travers l'observation de situations naturelles ou arrangées comme faisant partie des évaluations externes. Murray a indiqué que : « les observations peuvent être celles de novices ou d'experts et peuvent être destinées à l'identification des phénomènes de comportement ou des phénomènes cognitifs impliqués » (Murray, 1993).

3.2.15 Tâches structurées et classification quantitative de phénomènes

Cette approche se concentre sur une collection de données en limitant ou en organisant les réponses du sujet sous une recherche prospective. Les données sont recueillies en arrangeant ou en structurant les situations dans des évaluations externes, par exemple par des entrevues, des questionnaires et des aperçus. L'avantage de cette approche est que des données plus volumineuses et plus précises peuvent être obtenues en utilisant des tests plus structurés.

3.2.16 Études comparatives

C'est une classe de méthodes qui note les différences et les similitudes entre le comportement ou la conception d'un STI avec un standard ou un autre système. Ces méthodes sont intéressantes pour l'évaluation du système dans son entier et pour mesurer la performance de ce dernier dans une recherche prospective (Iqbal A. et al., 1999). Lewis Johnson a fait une étude comparative du "Tactical language training System" avec un groupe représentatif d'apprenants dans le but d'identifier la composante du logiciel qui améliore l'apprentissage (Lewis Johnson et Carole, 2005). Parallèlement Murray (1993) a suggéré trois méthodes pour réaliser de telles comparaisons : juger la performance du système avec un cas à succès bien connu (le standard de référence), ensuite argumenter le niveau théorique pour la généralité et

l'extensibilité d'une nouvelle approche par rapport aux autres systèmes (i.e corroboration théorique), et pour finir simuler le comportement d'un autre système (corroboration empirique ou duplication).

En somme, on peut dire qu'il existe plusieurs méthodes d'évaluation, le choix d'une méthode est fonction du problème à résoudre.

3.3 Une classification des méthodes

Le fait qu'il existe une grande variété de méthodes d'évaluation apporte une difficulté qu'on ne peut négliger lors d'une évaluation. Dans cette section nous allons voir une classification nous permettant de faire un choix éclairé des différentes méthodes d'évaluation.

Lors de l'établissement des conditions d'une évaluation, les éducateurs chargés d'évaluer le système sont confrontés à deux questions importantes à savoir : est-ce que c'est le système dans son ensemble qui est évalué ou juste une de ses composantes? Y a-t-il suffisamment d'apprenants ou d'autres conditions suffisamment acceptables pour faire une évaluation basée sur une évaluation expérimentale ?

La réponse à ces deux questions a permis à Iqbal et al. (1999) d'élaborer une charte classifiant les méthodes d'évaluation en deux dimensions. Dans un premier temps, une méthode est différenciée des autres par une échelle délimitant une évaluation externe (on considère le système dans son ensemble) et une évaluation interne (on teste les composantes internes du Système). Deuxièmement, une méthode est classifiée selon qu'elle appartient à la recherche expérimentale versus la recherche exploratrice. Dans le cas de la recherche exploratrice les méthodes doivent être réalisables avec un faible échantillon de données et dans un contexte inédit. Les méthodes d'évaluation qui sont classées comme expérimentales impliquent une manipulation des variables et nécessitent un groupe statistiquement important.

La figure 3.3.1 ci-après montre toutes les méthodes précédemment citées par rapport aux critères énoncés par Iqba et al. (1999). Les méthodes groupées dans le carré en bas à gauche de la charte sont des méthodes intéressantes pour l'évaluation des composantes de STI et ne nécessitent ni un grand échantillon ni une analyse statistique rigoureuse. Les méthodes nécessitant des groupes de contrôles d'études et destinées à l'évaluation des composantes sont groupées en bas à droite dans la charte. La partie haute de la charte montre les méthodes qui évaluent la performance globale du système. Les méthodes à droite nécessitent des études statistiques tandis que les méthodes à gauche n'en nécessitent point. Le point saillant de leur représentation est que dans le cas de la recherche expérimentale il existe des méthodes qui peuvent être utilisées à la fois pour l'évaluation de composantes et pour l'évaluation globale du système. La méthode 5 (l'analyse de sensibilité) en est une illustration. De même dans la recherche exploratrice, il existe des méthodes qui peuvent être utilisées à la fois pour l'évaluation interne et externe du système. Les méthodes 4 (instruction qualité/feed-back) et 13 (évaluations basées sur critère) en sont des exemples.

En général, on peut dire que cette classification proposée par Iqbal et al. bien que rigoureuse n'a d'intérêt que lorsque le STI est déjà totalement développé. Cependant dans le marché la plupart des STI sont à leur état prototype. Conséquemment comme les auteurs A. Iqbal et al. le font remarquer il serait bon d'ajouter à leur classification la dimension formative versus sommative. Cette dimension permettrait aux éducateurs d'évaluer les prototypes pour la conformité au curriculum, tout en gardant la classification ouverte pour l'évaluation sommative de tout produit entièrement développé. Cette classification est également limitée par le fait qu'elle ne classe pas les méthodes d'évaluation en fonction des composantes de l'architecture générale des STI (Iqbal A. et al., 1999).

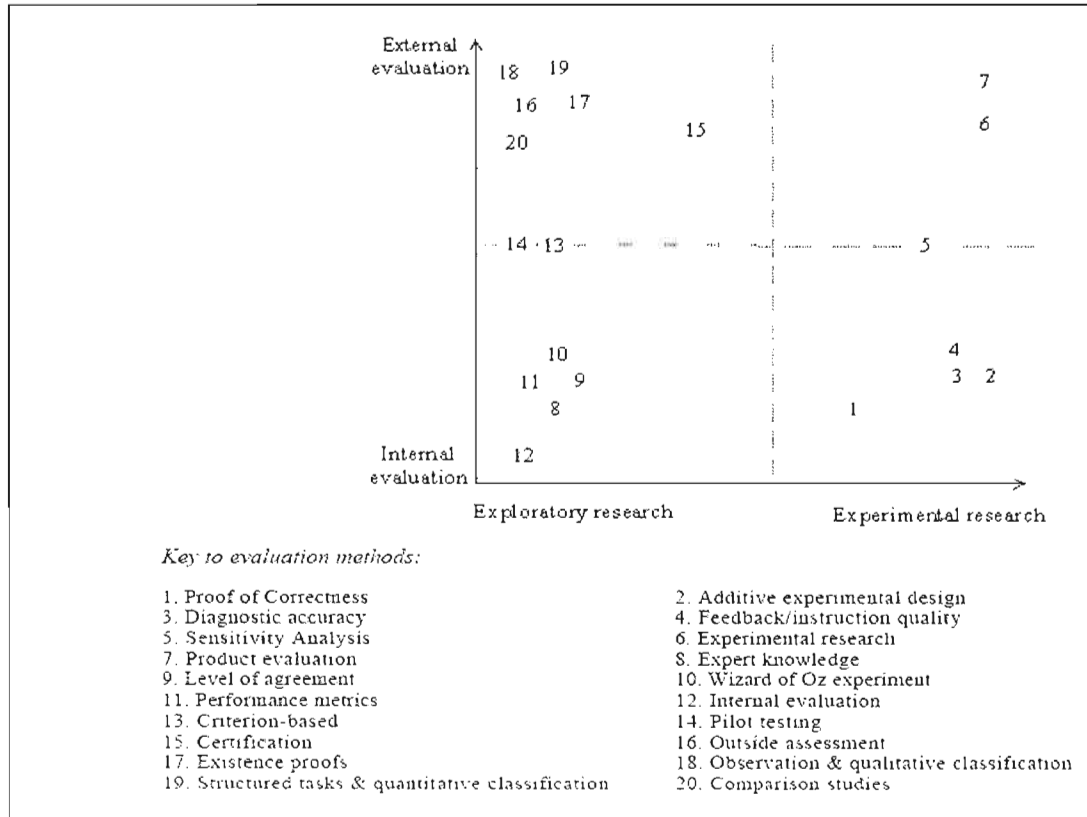


Figure 3.3.1 Charte de classification des méthodes d'évaluation (Iqbal A. *et al.*, 1999)

3.4 Points de vue sur l'évaluation des STI

La documentation en vogue nous apprend que l'évaluation d'un STI peut se faire pendant sa conception ou en fin de conception et peut se poursuivre à postériori. L'évaluation à postériori est celle qui survient lorsque le STI est complètement réalisé et en cours d'utilisation. Toutefois bien que l'on sache quand est-ce qu'il faut faire une évaluation, il reste encore à déterminer quels sont les critères de base pour réaliser une bonne évaluation. Le but de cette section est de présenter deux points de vue intéressants à considérer pour une bonne évaluation d'un STI.

3.4.1 Point de vue architectural

Les STI sont des systèmes complexes et de ce fait l'approche de l'évaluation de ses composantes varie en fonction de la composante à évaluer. Ainsi, il est important de

considérer l'adéquation de la technique d'évaluation par rapport aux composantes de l'architecture du STI. L'évaluation des composantes peut se faire de façon formative (durant la conception du STI) ou de façon sommative (lorsque le STI a été réalisé). Mark et Greer (1993) ont groupé plusieurs méthodes d'évaluation en tenant compte de l'architecture générale des STI à six composantes de McCalla et Greer (1990).

3.4.1.1 Base de connaissance

Elle concerne le stockage, la manipulation des connaissances du domaine et les raisonnements y afférents. Idéalement la précision de cette composante doit être vérifiée avant la complétude du système et assumée après. Ainsi, l'évaluation formative de cette composante doit être de première importance pour les développeurs. Tandis que l'évaluation sommative de la précision n'est intéressante que lorsque le domaine est imprécis.

Les techniques formatives d'évaluation de la base de connaissance sont : l'inspection des experts et le « *Turing test* ». En effet, il est facile de vérifier les faits et d'examiner les mécanismes de raisonnement dans les domaines moins complexes et bien compris. De même, si la base de connaissance du domaine est bien comprise ou clairement définie par des normes d'assurance, alors il est possible de déterminer si sa représentation est conforme à la norme. Cependant si le domaine n'est pas bien compris ou s'il est difficile de développer des études de cas ou d'autres mesures pour déterminer la précision du domaine, ou si la représentation des connaissances est difficile à inspecter, l'inspection des experts peut ne pas être désirable. Si l'humain peut être considéré comme un bon standard de comparaison pour l'expertise du domaine alors le Turing test peut être une évaluation adéquate (Mark et Greer, 1993).

3.4.1.2 Le modèle pédagogique

Le modèle pédagogique utilise des connaissances sur la façon d'enseigner. En principe, le modèle pédagogique d'un STI devrait être une composante qui conçoit et dirige une session en accord avec les méthodes d'enseignement et leurs conditions d'utilisation. Un objectif des développeurs des STI est de construire un système qui crée un enseignement personnalisé en accord avec les caractéristiques des apprenants tels leurs connaissances de base, leur niveau

de développement cognitif, leurs styles d'apprentissage ceci en vue d'améliorer l'efficacité de l'apprentissage de l'apprenant.

Les standards auxquels le modèle pédagogique peut être comparé sont les théories d'enseignement et les tuteurs humains. L'enseignement n'est pas nécessairement bien compris ou explicitement décrit, ce qui le rend difficile à évaluer. Les techniques d'évaluation formatives comme l'observation du comportement de surface de l'instruction assistée par ordinateur (CAI) et l'utilisation de liste de vérification s'appliquent mal à l'évaluation du modèle pédagogique des STI.

Si les standards pour évaluer la pertinence du modèle pédagogique d'un STI était développés, ils devraient refléter les critères de la NSTA⁷(Klopfer, 1986)qui inclut des considérations spécifiques comme : l'intervalle des méthodes d'enseignement offertes par un programme, le niveau d'adaptation d'un programme aux différences individuelles des apprenants, et les problèmes plus généraux (tel le degré avec lequel l'instruction s'appuie sur la recherche éducative et psychologique de l'enseignement). La manière dont de tels critères peuvent être évalués de façon formative ou sommative demeure encore inconnue.

Toutefois, l'analyse de sensibilité et la certification peuvent être des approches possibles d'investigation. Le « *Turing test* » et les techniques expérimentales sont difficilement applicables au comportement isolé d'une composante. Le comportement du modèle pédagogique dépend de l'apprenant et du modèle de connaissance. Il est difficile de produire une composante réaliste d'enseignement libre de toutes sources de confusion avec les autres composantes. Une possibilité d'évaluation consisterait à comparer expérimentalement une composante d'un STI avec d'autres composantes identiques. Aleven, Popescu et Koedinger l'ont fait avec deux versions du Geometry Explanation tutor pour voir si le fait d'utiliser le langage naturel ou le menu de choix améliorerait l'apprentissage (Aleven, Popescu et

⁷ NSTA : National Science Teachers Association

Koedinger, 2003). Le résultat de leur évaluation était que l'utilisation du langage naturel était légèrement supérieure à l'utilisation du menu de choix (Mark et Greer, 1993).

3.4.1.3 Le modèle de l'apprenant

Le modèle de l'apprenant sert à diagnostiquer et modéliser la compréhension de l'apprenant. Le diagnostic sert à examiner le comportement de l'apprenant pour en obtenir des informations significatives. La modélisation lit l'historique de l'apprenant de façon caractéristique et maintient ses informations pour l'usage du système. Le standard auquel le modèle apprenant d'un STI peut être comparé est l'apprenant lui-même, dont les connaissances peuvent être changeantes, contradictoires et difficiles à observer.

Les problèmes de validité, de fiabilité, d'objectivité et de standardisation peuvent être liés aussi au modèle apprenant. Le modèle de l'apprenant est valide dans la mesure où il reflète correctement l'apprenant dans le temps. En cela, le modèle de l'apprenant diffère de la plupart des instruments de tests, qui ne traitent pas des changements temporels exceptés dans le sens qu'ils peuvent être employés en tant que mesures discrètes de l'apprenant à des intervalles sur une certaine période de temps. Puisque que le modèle apprenant est basé sur l'idée de la surveillance des changements dans le comportement de l'apprenant et leurs compréhensions dans le temps, le critère de la fiabilité s'applique mal au modèle apprenant.

La validation du modèle apprenant est une considération importante pendant l'évaluation formative et sommative. Après que la validité ait été établie, les développeurs pourront s'interroger sur la fiabilité pendant l'évaluation sommative. Validité et fiabilité sont plus efficacement évaluées à travers des études expérimentales. Pour déterminer la validité d'un diagnostic, l'information diagnostiquée obtenue par le système est comparée à l'information diagnostiquée obtenue indépendamment. De la même façon, un modèle apprenant peut être validé en vérifiant si les mesures répétées du modèle d'apprenant sont uniformément conformes aux mesures indépendantes dans le temps. Pour évaluer la fiabilité, l'information des sessions comparables doit être examinée pour s'assurer de la consistance des résultats produits par le système (Mark M. A. et Greer J. E, 1993).

3.4.1.4 Composante de communication

La composante de communication (interface) présente l'information à l'apprenant et récupère les réponses de ce dernier. Mark et Greer (1993) font remarquer que si on soustrait les problèmes de qualité conceptuelle et génie logiciel, un STI ne sera d'aucune utilité si l'apprenant n'est pas capable de comprendre l'information qui lui est présentée. Pour évaluer cette composante, l'utilisation de normes (IHM) ou liste de critères peut être utile au développement de cette composante. Le test pilote ne servirait qu'à identifier les problèmes de l'interface. Une autre approche adaptable à l'évaluation formative ou sommative de cette composante est l'étude comparative de deux interfaces pour déterminer laquelle est plus efficiente (Mark M. A. et Greer J. E, 1993).

3.4.1.5 Le modèle adaptatif

Un STI peut être capable d'adapter son comportement avec sa composante d'apprentissage. Toutefois, si un STI peut modifier son approche pédagogique à l'image d'un enseignant humain en se servant de ses propres informations sur les apprenants, cela aura des conséquences sur l'évaluation du STI. Conséquemment, l'évaluation du système en son entier et/ou des composantes directement ou indirectement affectées par le modèle d'apprentissage ne sera valide qu'au moment où elle a été effectuée. De ce fait, il est essentiel, pour les STI possédant cette composante (modèle d'apprentissage), d'évaluer le changement de leur comportement dans le temps. En d'autres termes, lorsque cette composante est évaluée, il faut tenir compte de son impact autant sur l'apprentissage de l'apprenant que sur le fonctionnement des autres composantes du STI.

Un cadre pour l'évaluation des systèmes hypermédia adaptatifs (SHA) a récemment été proposé par Gupta (Gupta et Grover, 2004). Ce cadre a la particularité de traiter l'évaluation comme faisant partie intégrante du processus de développement et tient compte de la connexion internet. Il y a quatre dimensions orthogonales entre les différentes composantes : l'environnement (le contexte dans lequel évolue le SHA), l'adaptation (le type d'adaptation utilisée), le processus de développement (les étapes du cycle de vie génie logiciel utilisé pour développer le SHA) et les modèles d'évaluations (les couches du SHA qui ont été évaluées dans le contexte des quatre autres dimensions). La figure 3.4.1.5 ci-après est une illustration

du dit cadre d'évaluation. Une adaptation de ce cadre pourrait aider à l'évaluation des systèmes tutoriels intelligents adaptatifs (Mark et Greer, 1993).

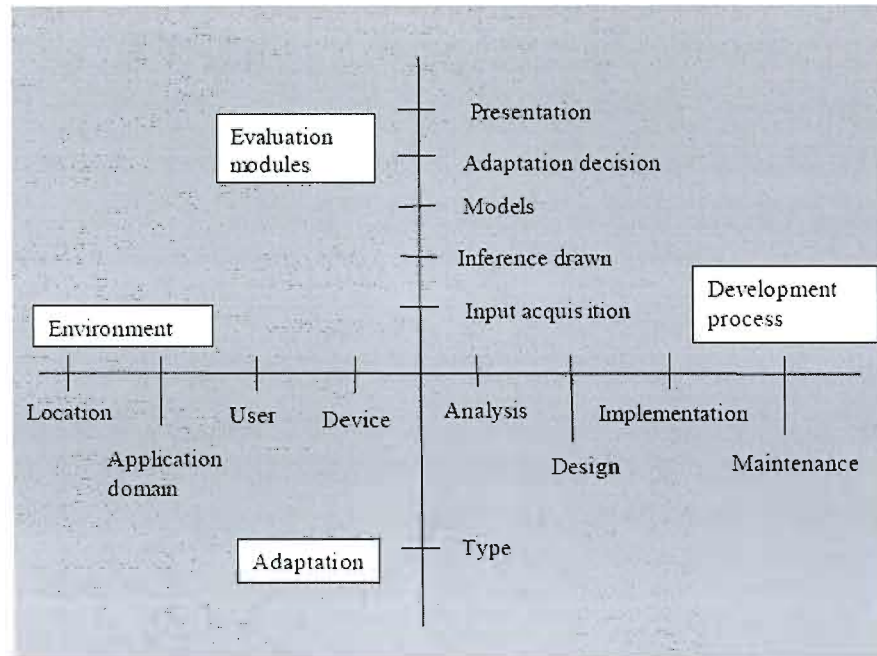


Figure 3.4.1.5.1 Cadre pour l'évaluation des SHA

3.4.1.6 Système de contrôle

L'ensemble des opérations d'un STI peut être géré par une composante de contrôle. De ce fait, cette composante a d'importantes implications sur le comportement du système en termes d'interaction et de vitesse. L'évaluation appropriée de cette composante est fonction de son implémentation. Puisque la composante de contrôle est concernée par le comportement des dispositifs fondamentaux du système tels que la planification des processus de différents composants, les techniques conventionnelles d'évaluation de performances d'ordinateur pourraient elles aussi être appropriées pour ce composant. Une telle analyse est le plus susceptible d'avoir un intérêt pour les experts formateurs concernés par l'amélioration de l'exécution du système. Malgré tout, les experts de l'évaluation

sommative sont plus concernés par les aspects éducatifs d'un STI que par sa performance informatique (Mark et Greer, 1993).

3.4.2 Utilisabilité-utilité-acceptabilité

Un système informatique peut être évalué sur plusieurs aspects : pour son adéquation aux objectifs de haut niveau du client (utilité), sa capacité à permettre à l'apprenant d'atteindre facilement ses objectifs (utilisabilité) et sa compatibilité avec les valeurs du client. Ces critères peuvent être évalués de façon empirique et/ou par inspection comme le signale Tricot (Tricot et al., 2003). L'inspection est réalisée par un "expert", qui applique de façon plus ou moins explicite des critères d'évaluation. L'évaluation empirique, quant à elle, consiste à interpréter les performances des apprenants à qui l'on prescrit une tâche, et plus généralement à interpréter leurs comportements, attitudes, et opinions. En remplaçant l'apprenant par l'apprenant, ces critères peuvent être utilisés pour l'évaluation des STI.

3.4.2.1 L'évaluation de l'utilité

Rappelons que classiquement l'évaluation de l'utilité consiste à observer si l'utilisateur est capable d'accomplir sa tâche à partir des fonctionnalités du système. Dans le cadre des STI l'objectif à atteindre est découpé en deux niveaux à savoir : l'apprentissage (de la discipline enseignée et non du système) et la résolution des problèmes par le système. Le critère d'utilité pour les STI qu'il nous importe de mesurer est non la résolution de problèmes, mais l'apprentissage (Nogry, Jean-Daubias et Ollagagnier-Beldame, 2004). En effet, la difficulté ou l'impossibilité de résoudre un problème peut faire partie de la pédagogie du système. Par conséquent, évaluer ce critère revient à évaluer la pédagogie du système. Diverses méthodes d'évaluation de l'apprentissage existent. Ces méthodes, issues de techniques non spécifiques aux STI, ont été adaptées pour évaluer de tels systèmes.

D'une part, en ce qui concerne l'évaluation empirique, dont le principe général repose sur la comparaison de groupe sujet soumis à des tâches, les techniques d'évaluation peuvent être groupées en deux catégories : les méthodes quantitatives, objectives, qui permettent de

quantifier l'apprentissage et donnent des résultats généralisables (les méthodes comparatives), mais qui ne permettent pas de prendre en compte des situations complexes, et des méthodes qualitatives (verbalisation, collecte de documents, méthodes ethnographiques, etc...) qui permettent de prendre en compte la complexité de la situation, mais dont les résultats n'ont pas vocation à être généralisés, leur validité étant restreinte à des contextes proches de la situation étudiée. Bien que nous utilisions cette distinction entre méthodes qualitatives et quantitatives, ces deux types d'approches forment un continuum (Nogry, Jean-Daubias et Ollagagnier-Beldame, 2004).

D'autre part, l'évaluation par inspection de l'utilité vise à s'assurer de l'adéquation de la méthodologie pédagogique. À cet effet, Tricot a retenu sept critères présentés dans le tableau 1. Cependant comme il faudrait remarquer qu'un STI est intégré dans un cadre d'enseignement ou de formation, la situation peut prendre en charge certains aspects. De ce fait, certains critères peuvent être satisfaits par le contexte et non par le STI lui-même (Tricot *et al.*, 2003).

Type d'évaluation /Dimensions	Empirique	Par inspection
Utilité	Adéquation entre objectif défini et apprentissage effectif Adéquation entre dispositif et format de la connaissance à acquérir Différence entre niveau de connaissances initial et terminal Mesures par des tâches de . <ul style="list-style-type: none"> • reconnaissance • rappel (contenu / structure) • résolution de problème • détection d'erreurs • production 	Précision et présentation des objectifs Adéquation contenus /objectifs Précision du scénario didactique Adéquation scénario / objectifs / contenus Mise en œuvre des processus cognitifs et méta cognitifs Régulation Évaluation

Tableau 3.4.2.1.1 Présentation des mesures et critères pour l'évaluation de l'utilité d'un STI (Tricot *et al.*, 2003)

3.4.2.2 Évaluation de l'utilisabilité

ISO 9241-11 stipule qu'un logiciel est utilisable lorsque l'utilisateur peut réaliser sa tâche (efficacité), qu'il consomme un minimum de ressource (efficience) et que le système est agréable à utiliser (satisfaction de l'utilisateur). Dans le cas particulier des STI, cette catégorie de critères répond à la question : le STI est-il aisé à prendre en main, à utiliser, sans perdre de temps et sans faire d'erreur de manipulation? En conséquence, l'utilisabilité d'un STI se joue au niveau de son interface (cohérence, lisibilité, etc.), de sa navigation (simplicité, exhaustivité des déplacements possible, etc..) et de sa cohérence avec l'objectif et le scénario didactique (Nogry, Jean-Daubias et Ollagagnier-Beldame, 2004).

Les critères d'évaluation de l'utilisabilité d'IHM doivent être adaptés aux STI. De nombreux travaux ont été réalisés dans l'adaptation de ces derniers (Jean, 2000). Par ailleurs, concernant l'évaluation par inspection, les critères de Bastien et Scapin (1993) sont souvent cités dans le domaine de l'évaluation des IHM. L'efficacité de ces derniers a été évaluée et comparée avec d'autres critères comme la norme ISO 9241-11 (Bastien, Scapin et Leulier, 1999.). Par exemple, l'application du critère de gestion des erreurs (Bastien et Scapin, 1993) aux STI nécessite de faire la distinction entre deux types d'erreurs : les erreurs dans l'utilisation du logiciel et les erreurs au sens de réponses erronées (erreurs conceptuelles, (Jean, 2000)). Si les erreurs d'utilisation du logiciel doivent être empêchées ou corrigées, suivant ainsi la recommandation de Bastien et Scapin, la gestion des erreurs conceptuelles doit être traitée indépendamment en fonction de la théorie cognitive sous-jacente au STI. Ainsi, dans une perspective behavioriste on choisira de corriger immédiatement chaque erreur, tandis que dans d'autres perspectives on pourra laisser l'apprenant se rendre compte lui-même de ses erreurs.

En ce qui concerne l'évaluation empirique, un exemple célèbre est celui de Nielsen (Nielsen, 1993). Cette méthode consiste à réaliser des tests sur l'utilisateur (observation du comportement et suivi des apprenants dans une tâche prescrite) et des entretiens avec ces mêmes utilisateurs. L'analyse de parcours réalisé par chaque utilisateur (les éléments du STI traités, dans quel ordre, le temps passé sur chaque élément) complète l'analyse empirique de ce critère (Tricot et al., 2003). Nogry et al. (2004) trouvent que les méthodes d'évaluation

empiriques sont plus flexibles et semblent plus appropriées aux STI dans la mesure où les utilisateurs observés sont représentatifs des apprenants qui utiliseront ensuite le système. Les méthodes classiques d'observations et d'entretiens peuvent donc être utilisées. Toutefois, ces observations doivent être effectuées dans le contexte d'utilisation envisagé pour le logiciel afin de déterminer si les apprenants ont le comportement et les résultats escomptés dans des conditions proches de la réalité.

Type d'évaluation /Dimensions	Empirique	Par inspection
Utilisabilité	Possibilité d'apprendre à utiliser le système Gestion et Prévention des erreurs Mémorisation du fonctionnement Efficience Sentiment de satisfaction Évaluation par <ul style="list-style-type: none"> • observations • entretiens • analyse des parcours 	Guidage et Incitation Groupement / Distinction des items par localisation ou format Feed-back immédiat et nature du feed-back Charge de travail Contrôle explicite Adaptabilité Gestion des erreurs Qualité des messages Homogénéité et cohérence Signification des codes et dénominations

Tableau 3.4.2.2 Présentation des mesures et critères pour l'évaluation de l'utilisabilité d'un STI (Tricot et al., 2003)

3.4.2.3 Évaluation de l'acceptabilité

Dans le cadre des STI, l'acceptabilité est définie par Tricot et al. (2003) comme la valeur de la représentation mentale (attitudes, opinions plus ou moins positives) de l'utilité et de l'utilisabilité du système. En d'autres termes, l'acceptabilité concerne la décision d'utiliser le STI. Cette catégorie de critères répond à la question : le STI est-il compatible avec les valeurs, la culture des utilisateurs et l'organisation dans lesquelles on veut l'insérer? Résultat, les mesures empiriques de l'évaluation de l'acceptabilité relèvent de champs très divers. Les

critères d'évaluation empirique et par inspection de l'accessibilité sont résumés dans le tableau 3.4.2.3 ci-après.

Amiel et al. ont montré dans leur étude que les représentations de l'utilité et de l'utilisabilité d'un STI pouvaient être les variables les plus prédictives de la décision concernant son utilisation ou prescription (Amiel et al., 2002). En complément à cela, Tricot et al. (2003) affirment que toute relation entre ces trois dimensions (utilité, utilisabilité, acceptabilité) complémentaires est pertinente et suffisante à une évaluation des STI.

Type d'évaluation /Dimensions	Empirique	Par inspection
Acceptabilité	Motivation Affects Culture Valeurs Évaluation par : <ul style="list-style-type: none"> • observations • entretiens • questionnaires 	Acceptabilité en termes d'adéquation aux : Besoins ou objectifs de l'institution Attentes des apprenants Caractéristiques des apprenants Acceptabilité en termes de compatibilité avec : L'organisation du temps L'organisation des lieux Présence du matériel nécessaire Planification et suivis lisibles et cohérents Visibilité des résultats

Tableau 3.4.2.3 Présentation des mesures et critères pour l'évaluation de l'accessibilité d'un STI (Tricot et al., 2003)

3.4.3 Génie logiciel

Les STI étant pluridisciplinaires voire transdisciplinaires, la notion d'« ingénierie des STI » s'étend au-delà du point de vue technologique, en l'occurrence informatique. Ainsi, Pierre Tchounikine a défini comme ingénierie des STI les travaux visant à définir des concepts, méthodes et technique reproductibles et/ou réutilisables facilitant la mise en place (conception – réalisation – expérimentation – évaluation – diffusion) d'environnement de

formation ou d'apprentissage (dans leur articulation avec les dispositifs informatiques d'aujourd'hui) en permettant de dépasser le traitement ad hoc des problèmes. En d'autres termes, pour faire face à la construction d'un STI répondant à des contraintes identifiées (domaine, niveau, contexte, objectif pédagogique, etc.), l'équipe de conception pourrait faire appel à un ensemble de connaissances (et éventuellement d'outils) capitalisées (« éléments de méthodes et de techniques ») qui guideront le processus de conception, de réalisation, d'évaluation et d'expérimentation du système. Ceci sous-entend que cet ensemble de connaissances aura été capitalisé (identifié, décrit) car reconnu comme utile dans un spectre d'application donné (Tchounikine et al., 2002).

Il n'existe pas de normes de référence sur l'élaboration d'un STI. Toutefois, Tchounikine avec Platon-1 a proposé un cadre d'analyse pertinent, permettant de guider le développement d'un STI. Platon-1 n'est pas un cadre normatif, mais un ensemble de dimension d'analyse et de compréhension de travaux menés sur le domaine qu'il peut être pertinent de préciser ou de compléter selon le contexte (Tchounikine et al., 2004). Les développeurs (les programmeurs) de STI contrairement aux analystes bénéficient de nombreux outils d'aide au développement.

En effet, les progrès techniques actuels de l'informatique pertinents pour les STI (les hypertextes, les systèmes à base de connaissance, les TIC aujourd'hui) a conduit à des travaux dont le principe est de construire une « coquille vide » correspondant à cette nouvelle technologie et à proposer cette coquille comme un outil permettant de « générer » des STI. Cette approche, plus fondée sur les possibilités technologiques que sur les spécificités de l'apprentissage avec des machines, a conduit à la construction de dispositifs génériques (proposés par les promoteurs de la technologie concernée plus que par les spécialistes de STI) mais à peu de systèmes effectivement utilisés (Tchounikine et al., 2002).

De plus, ces progrès technologiques, notamment l'utilisation accrue de l'internet a entraîné la multiplication croissante des ressources d'apprentissage. En vue de faciliter l'indexation et la recherche de ces derniers, des normes et standards tels LOM, SCORM et IMS ont été proposés. Ces normes permettent également l'évaluation de l'utilisabilité de plateforme web destinée à l'apprentissage. Bien que des repères permettant la conception et la réutilisation

des STI existent, l'évaluation de ces derniers reste très problématique du point de vue de l'ingénierie.

La principale difficulté de l'évaluation d'un STI vient de l'objet central que l'on doit considérer, c'est-à-dire l'apprentissage. En effet, un STI a pour objectif principal soit un apprentissage en tant que tel (apprentissage d'un concept d'un domaine donné ou apprentissage d'une compétence à travers une activité dans un domaine), soit différents aspects facilitant l'apprentissage (modification de l'appréhension d'un problème, éveil d'un intérêt, création de liens sociaux, etc.). Dans les deux cas, la réponse concernant l'atteinte de l'objectif n'est pas booléenne, et les facteurs pouvant influencer l'atteinte de l'objectif sont difficilement identifiables. Cependant, l'évaluation des « produits » est un aspect clé de la mise en place d'une ingénierie. Le fait que le logiciel réponde à sa spécification informatique permet de valider le versant « mise en œuvre » informatique, mais en aucun cas le STI en tant que tel. De ce fait, l'élaboration d'un cadre de référence sur l'évaluation des STI serait un outil précieux à l'ingénierie de ces derniers. Ce cadre se devra d'insister tant sur la notion même d'évaluation (ce que l'on peut/doit évaluer et pourquoi) que sur les procédés (Tchounikine et al., 2002).

3.5 Perspectives

Les raisons justifiant l'évaluation d'un STI de nos jours sont multiples et touchent les trois aspects suivants : scientifique (apprendre des erreurs, évolution STI,...), éducatif (améliorer l'apprentissage) et économique. Du fait de leur nature complexe, l'évaluation des STI n'est pas une tâche aisée.

Ainsi, nous avons constaté qu'il existe différentes mesures communes d'évaluation de l'apprentissage. Cette dernière (l'apprentissage) étant un point central dans l'évaluation des STI. Ces mesures seront rarement utilisées isolément lors de l'évaluation d'un STI. Par ailleurs, plusieurs techniques peuvent être utilisées lors de l'évaluation d'un STI. Les particularités du critère à évaluer doivent être prises en compte dans le choix de la technique

à utiliser. A. Iqbal a proposé une classification à deux dimensions de ces techniques (Iqbal A. *et al.*, 1999). Dans un premier temps, une méthode est différenciée des autres par une échelle délimitant une évaluation externe et une évaluation interne. Deuxièmement, une méthode est classifiée selon qu'elle appartient à la recherche expérimentale versus la recherche exploratrice. L'ajout d'une dimension qui classerait les méthodes selon qu'on évalue un logiciel achevé ou en cours de réalisation aiderait les développeurs et les évaluateurs de systèmes dans le choix de techniques d'évaluation. Nous avons ensuite examiné trois points de vue complémentaires pouvant nous aider à bien évaluer les STI. Du point de vue architectural, il est important d'évaluer chaque composante du système. Toutefois avec ce point de vue, nous avons remarqué qu'il était difficile d'évaluer la composante adaptative. Par ailleurs l'utilité, l'utilisabilité et l'acceptance sont également pertinents pour l'évaluation d'un STI. En effet, ces trois dimensions interreliées nous permettent de tenir compte de l'impact qu'a le système sur l'apprenant. En définitive, ces deux points de vue permettent de préciser l'idée de qualité qui est nécessaire à l'élaboration d'un cadre d'évaluation d'un STI qui est un problème majeur dans l'ingénierie des STI.

CHAPITRE IV

PROPOSITION D'UNE GRILLE D'ÉVALUATION DES SYSTÈMES TUTORIELS INTELLIGENTS

Dans le présent chapitre, nous allons proposer une grille d'évaluation des STI. Pour ce faire, nous allons commencer par justifier la logique de conception de notre grille. Nous terminerons en présentant les composantes ergonomiques et utilitaires de notre grille d'évaluation.

4.1 Justification et objectif de notre grille d'évaluation

Nous avons remarqué dans le chapitre précédent que l'ergonomie et l'utilité d'un STI sont des facteurs déterminants de son acceptabilité. De ce fait, ces deux facteurs constitueront les deux axes de la grille d'évaluation que nous allons élaborer dans le présent chapitre. De plus, afin que notre évaluation soit très ciblée nous utiliserons une approche d'évaluation par modèle comme suggéré par Marc et Greer (1993). Ce modèle adaptatif permettrait au STI d'adapter automatiquement sa pédagogie voire la changer en fonction du contexte. Toutefois, étant donné que la littérature ne nous a pas présenté de STI adaptatif, cette composante ne sera pas traitée par notre grille d'évaluation.

L'aspect ergonomique permet à un individu de se familiariser plus facilement avec un outil ou de le rejeter. Ce facteur s'adresse au STI dans son ensemble, mais plus particulièrement à son interface de communication. Pour juger de ce facteur dans notre grille, nous allons nous

son interface de communication. Pour juger de ce facteur dans notre grille, nous allons nous servir des critères de Bastien et Scapin (1993) tels que le recommande la littérature (voir section 3.4.2.2 Évaluation de l'utilisabilité).

L'aspect utilitaire quant à lui concerne plus la capacité du système à permettre à l'apprenant de développer voire d'acquérir ses habiletés ou connaissances. En d'autres termes, ce facteur concerne la pédagogie du système. Le côté pédagogique d'un STI est lié à l'interaction entre le modèle expert, la base de connaissance et le modèle pédagogique. Comme nous l'avons vu avec Marc et Greer (1993), l'évaluation du modèle pédagogique est une chose complexe. Toutefois, Tricot et al. (2003) ont recensé des critères d'évaluation de l'utilité d'un STI (voir section 3.4.2.1). Ces critères d'évaluation nous serviront de points de repère pour l'évaluation de l'utilité d'un STI.

Notre grille a pour but de permettre au client d'un STI d'avoir une vision claire et précise des capacités du STI évalué. Ainsi, nous avons groupé les questions suivant les critères d'ergonomie ou de l'utilité. De plus, afin de diminuer les problèmes d'ambiguïté des réponses aux questions de notre grille nous avons fixé un barème de notation à trois points à savoir:

- 0 point : Le STI ne répond pas de façon satisfaisante à l'exigence;
- 1 point: le STI satisfait partiellement à l'exigence;
- 2 points: le STI satisfait pleinement à l'exigence;
- X- : l'exigence est inapplicable au STI évalué.

Enfin, les questions de notre grille d'évaluation sont non exhaustive, mais offrent une vision suffisamment complète du STI évalué. Un utilisateur de notre grille a donc la possibilité d'y ajouter des questions afin de mieux l'adapter à l'évaluation qu'il souhaite mener.

Dans les sections qui vont suivre, nous présenterons comment à partir des critères de Bastien et Scapin nous contruirons une grille d'évaluation ergonomique d'un STI. Par la suite, à

partir des critères recensés par Tricotn nous déduirons le questionnaire pour l'évaluation des quatre principales composantes concernées par l'évaluation de l'utilité d'un STI. Enfin, en annexe, nous présenterons notre grille d'évaluation dans son intégralité.

4.2 Grille d'évaluation de la dimension ergonomique

La dimension ergonomique est très pertinente pour l'évaluation d'un STI car il faut que l'apprenant soit à l'aise avec l'utilisation du système sinon il va le rejeter. Pour l'évaluation de cette dimension, nous avons retenu les critères de Bastien et Scapin recommandés par la littérature. Afin de ne pas réinventer la roue pour élaborer les questions qui nous permettront de déterminer si un critère est satisfait ou non, nous nous sommes inspiré de la check-list de (Pierotti, 1996) que nous avons adaptée afin d'être conforme aux critères de Bastien et Scapin (1993) et à notre méthodologie d'évaluation. Par ailleurs dans les sous section suivante la description des critères est un résumé de la description général que font Bastien et Scapin (1997).

4.2.1 Guidage

Il concerne l'ensemble des moyens mis en œuvre pour conseiller, orienter, informer, et conduire l'utilisateur (dans notre cas l'apprenant) lors de ses interactions avec l'ordinateur (messages, alarmes, étiquettes, etc.). Le bon respect de ce critère facilite l'apprentissage et l'utilisation du système dans la mesure où il permet à l'apprenant :

- de savoir, à tout moment où il se trouve dans une séquence d'interactions, ou dans l'accomplissement d'une tâche ;
- de connaître les actions permises ainsi que leurs conséquences ;
- d'obtenir de l'information supplémentaire (Bastien et Scapin, 1997).

Ainsi, un bon guidage réduit les erreurs de l'apprenant et améliore les performances de ce dernier. Dans les sous-sections qui suivent nous allons voir les sous-critères qu'il implique.

4.2.1.1 Incitation

C'est l'ensemble de moyens et actions poussant l'apprenant à connaître les actions disponibles, les alternatives, l'état du contexte où il se trouve et les outils d'aide à sa disposition (Bastien et Scapin, 1997). Les questions qu'on peut se poser pour voir si ce sous-critère est respecté sont regroupées dans le tableau suivant :

Check-list	X	1	2	3	<u>Commentaires</u>
E.1.1.1 Le STI suggère à l'utilisateur des actions permettant de débiter la résolution d'un problème					
E.1.1.2 Le STI fournit l'option d'annulation de résolution d'une opération en cours					
E.1.1.3 Le STI fournit l'option de redémarrer la résolution d'un problème en cours					
E.1.1.4. Le STI fournit de l'aide à l'apprenant					
E.1.1.5 Le STI fournit des instructions sonores pertinentes pour guider l'apprenant					
E.1.1.6 Le STI fournit des instructions vidéo pertinentes pour guider l'apprenant dans la manipulation					
E.1.1.7 Le STI fournit des instructions textuelles pertinentes pour guider l'apprenant					
E.1.1.8 Le STI fournit un personnage animé attrayant et pertinent qui guide l'apprenant dans la résolution de problème					
E.1.1.9 Il existe une boîte de menu ou un message de feed-back indiquant les choix et le positionnement du curseur					
E.1.1.10 L'interface facilite toute sélection,					

lorsque possible					
E.1.1.11 L'interface facilite toute désélection, lorsque possible					

Tableau 4.2.1.1.1 Évaluation de l'ergonomie : l'incitation

4.2.1.2 Groupement/Distinction par le format ou le groupement

Ce critère concerne l'organisation visuelle des onglets d'information les uns par rapport aux autres. Il prend en compte la localisation et les caractéristiques graphiques (le format) pour indiquer les relations entre les divers items affichés, et leur appartenance ou non à une même classe d'items. De cette façon, l'apprentissage et la mémorisation des items ou groupe d'items s'en trouvent facilité (Bastien et Scapin, 1997).

Ainsi, les problèmes que pose ce critère sont les suivants :

Check-list	x	1	2	3	Commentaires
E.1.2.1 Le regroupement des items/menus/icônes matche avec le regroupement des tâches					
E.1.2.2 L'emplacement des items/icônes/menus est intuitif					
E.1.2.3 Les items fréquemment utilisés sont les plus accessibles					
E.1.2.4 Chaque item est un membre harmonieux d'une famille d'items					
E.1.2.5 Les instructions, les affichages et les messages d'erreurs apparaissent toujours aux mêmes endroits dans le menu					
E.1.2.6 Les informations sont présentées par ordre d'importance					

décroissante					
--------------	--	--	--	--	--

Tableau 4.2.1.2.1 Évaluation de l'ergonomie : le Groupement/Distinction par le format ou le groupement

4.2.1.3 Feedback Immédiat

Il concerne la capacité du système à fournir une réponse à l'apprenant, le renseignant sur l'action accomplie et sur son résultat. Tout ceci avec un délai approprié de réponse et une homogénéité selon les types de transactions (Bastien et Scapin, 1997). En d'autres termes, ce critère s'intéresse à la pertinence et au délai adéquat des rétro-contrôles (feedback). Le tableau ci-après contient les questions que nous avons recensées pour déterminer la satisfaction de ce critère par un STI.

Check-list	X	0	1	2	Commentaires
E.1.3.1 Il y a une sorte de feed-back pour chacune des opérations de l'apprenant					
E.1.3.2 Après la complétion d'une action (ou groupes d'actions) un feed-back indique que la prochaine étape d'actions peut débiter					
E.1.3.3 Il y a un feed-back visuel quand un objet est sélectionné ou déplacé					
E.1.3.4. Le temps de réponse du système est approprié aux tâches à exécuter					
E.1.3.5 Les messages de feed-back sont appropriés et suffisamment précis					

Tableau 4.2.1.3.1 Évaluation du Feedback Immédiat

4.2.1.4 Lisibilité

Il s'agit ici des caractéristiques matérielles de présentation des informations qui doivent en faciliter la lecture telle que : la typographie, le contraste avec l'arrière-plan, l'espacement

Bastien et Scapin, 1997). Le problème est donc de déterminer si les informations sont présentées dans un format facilitant la lecture. Le tableau ci-après présente la portion de notre grille d'évaluation concernant l'évaluation de ce sous-critère.

Check-list					Commentaires
E.1.4.1 Les informations sont présentées dans un format facilitant la lecture					
E.1.4.2 Le nom de chaque item est explicite des actions qui en découle					
E.1.4.3 Une icône sélectionnée est clairement visible quand elle est entourée d'autres icônes non sélectionnées					
E.1.4.4 Le format des items facilite leur lecture					
E.1.4.5 Dans un affichage multiple de pages d'écran de données, chaque écran est nommé de façon à décrire sa relation avec les autres					
E.1.4.6. Les couleurs sélectionnées correspondent aux espérances des codes de couleurs					
E.1.4.7 Pour les interfaces de questions/réponses, les questions sont claires et dans un langage simple					
E.1.4.8 Les menus sont plus en largeur (plusieurs options dans le menu) qu'en profondeur (plusieurs niveaux d'imbrication de menu)					
E.1.4.9 Les instructions en ligne apparaissent en des endroits pertinents sur le(s) écran(s)					

Tableau 4.2.1.4.1 Évaluation de l'ergonomie : la Lisibilité

4.2.2 Charge de travail

Ce critère touche tous les éléments de l'interface visant la réduction de la charge perceptuelle, mnémésique, et cognitive en vue de faciliter le dialogue apprenant-interface. En effet, la probabilité de faire des erreurs de manipulation est croît de pair avec la charge de travail. Ainsi, moins l'apprenant sera distrait par des informations non pertinentes plus il sera capable d'accomplir efficacement sa tâche. Par ailleurs, plus les actions requises seront courtes, plus rapides seront les interactions (Bastien et Scapin, 1997). En résumé, une charge de travail adéquate est caractérisée par une brièveté et une densité de l'information.

4.2.2.1 Brièveté

Ce critère vise la réduction maximale de la saisie et de la lecture d'information. Pour ce faire, il faudrait que :

- les éléments individuels d'entrée ou de sortie soient concis afin de réduire la charge mnémésique;
- le nombre d'action nécessaire à la réalisation d'une tâche soit réduit au maximum (Bastien et Scapin, 1997).

Les questions permettant de déterminer si un STI a répondu à ce sous-critère sont regroupées dans le tableau ci-après :

Check-list	X	0	1	2	<u>Commentaires</u>
E.2.1.1 Les détails excessifs de conception sont évités					
E.2.1.2 Les éléments d'entrée et sortie sont suffisamment concis					
E.2.1.3 La quantité de fenêtres à gérer en même temps est minimale					
E.2.1.4. L'interface utilise entre 4 à 7 couleurs toutes bien éloignées les unes des autres dans le spectre visible					
E.2.1.5 Les techniques pour attirer l'attention sont utilisées avec soin sans être excessives					
E.2.1.6 L'utilisation des raccourcis clavier et/ou de boutons de raccourcis permet de minimaliser la circulation à travers les menus					
E.2.1.7 Le nombre d'icônes est compris entre douze et vingt					

Tableau 4.2.2.1.1 Évaluation de l'ergonomie : brièveté

4.2.2.2 Densité informationnelle

Ce critère concerne la charge de travail de l'apprenant par rapport aux informations qui lui sont présentées dans leur ensemble du point de vue perceptuel et cognitif. Dans la plus part des cas, une charge de travail trop élevé est un diminue les performances de l'apprenant (Bastien et Scapin, 1997). Il est par conséquent important réduire la charge informationnelle au maximum. Dans le tableau ci-dessous sont regroupées les questions permettant de déterminer si ce sous-critère est satisfait par le STI

Check-list	X	0	1	2	<u>Commentaires</u>
E.2.2.1 Le regroupement des items est intuitif					
E.2.2.2 Le nom de chaque item est explicite des actions qui en découlent					
E.2.2.3 Tous les items à travers le système sont identifiés par un nom					
E.2.2.4 Chaque écran commence avec un titre ou un en-tête décrivant le contenu					
E.2.2.5 Une légende explicative des couleurs est fournie si les couleurs utilisées sont nombreuses ou difficilement compréhensibles					
E.2.2.6 Toutes les actions à travers le système ont des noms significatifs qui ne sont pas ambigus					
E.2.2.7 Les informations les plus importantes sont placées en début d'affichage lorsque cela est possible					
E.2.2.8 Pour les interfaces questions/réponses, les réponses valides sont listées.					
E.2.2.9 Le son est doux pour un feed-back positif et grave dans le cas contraire.					
E.2.2.10 Les noms de tous les menus sont judicieux à travers le système					

Tableau 4.2.2.2.1 Évaluation de l'ergonomie : Densité informationnelle
4.2.3 Contrôle Explicite

Ce critère s'intéresse à la fois à la prise en compte par le système des actions explicites et au contrôle qu'ont les apprenants sur le traitement de leurs actions. Le fait que les apprenants définissent et contrôlent explicitement leurs entrées a pour effet de limiter les erreurs. De

plus, le contrôle qu'ont les apprenants sur le dialogue est un facteur d'acceptation du système (Bastien et Scapin, 1997).

4.2.3.1 Actions Explicites

Il concerne la relation entre les actions de l'utilisateur et les réponses du système qui doit être explicite. De ce fait, le système doit permettre seulement les actions demandées par l'apprenant au moment où ce dernier le désire. Dans le tableau ci-dessous sont regroupées les questions permettant de déterminer si ce sous-critère est satisfait par le STI (Bastien et Scapin, 1997).

Check-list	X	0	1	2	<u>Commentaires</u>
E.3.1.1 Le temps de réponse est approprié à chaque tâche					
E.3.1.2 La vitesse du clavier, mouvement du curseur, sélection de la souris est de 50 à 150 millisecondes					
E.3.1.3 Le temps des tâches simples et fréquentes est de 1 seconde					
E.3.1.4. Le temps des tâches communes est de 2 à 4 secondes					
E.3.1.5 Le temps des tâches complexes est de 8 à 12 secondes					
E.3.1.6 Le temps de réponse est approprié au traitement cognitif de l'apprenant					
E.3.1.7 S'il y a un temps de traitement d'une durée observable (plus que cinquante secondes), l'apprenant est tenu					

informé des progrès du système					
--------------------------------	--	--	--	--	--

Tableau 4.2.3.1.1 Évaluation de l'ergonomie : Actions explicites

4.2.3.2 Contrôle de l'utilisateur

Il concerne la relation entre les actions de l'utilisateur et les réponses du système qui doit être explicite. De ce fait, le système doit permettre seulement les actions demandées par l'apprenant au moment où ce dernier le désire. Dans le tableau ci-dessous sont regroupées les questions permettant de déterminer si ce sous-critère est satisfait par le STI (Bastien et Scapin, 1997).

Check-list	X	0	1	2	Commentaires
E.3.2.1 Si le système utilise plusieurs couches de fenêtres, il est facile à l'apprenant des les réarranger sur l'écran					
E.3.2.2 Si le système utilise plusieurs couches de fenêtres, l'apprenant peut facilement se déplacer d'une fenêtre à l'autre					
E.3.2.3 Quand une tâche est terminée, le système attend une confirmation de l'apprenant					
E.3.2.4 L'apprenant peut annuler des opérations en cours					
E.3.2.5 L'apprenant doit confirmer l'exécution de commandes qui ont des conséquences drastiques					
E.3.2.6 L'apprenant peut réduire le temps d'entrée des données en copiant ou en modifiant des données déjà existantes					
E.3.2.7 Si le système a plusieurs niveaux de menus, il y a un mécanisme permettant le retour au menu précédent					
E.3.2.8 Si le système permet de reculer au menu précédent,					

l'apprenant peut modifier les choix précédemment effectués					
E.3.2.9 L'apprenant peut avancer et reculer à travers les champs ou boîtes de dialogues d'options					
E.3.2.10 Les touches/boutons de fonctions aux conséquences drastiques offrent un moyen d'annulation					
E.3.2.11 La continuité de pensée est nécessaire et les informations dont il faut se souvenir à travers les réponses datent de moins de deux secondes					

Tableau 4.2.3.2.1 Évaluation de l'ergonomie : contrôle de l'utilisateur

4.2.4 Adaptabilité

Ce critère concerne l'aptitude du système à réagir de façon contextuelle en accord avec les besoins et préférences des apprenants. Plus il y a d'alternatives pour exécuter une même tâche, plus les chances que l'apprenant puisse choisir et maîtriser l'une d'entre elles qui lui conviendraient le mieux. En raison de cela, Il est important de mettre à la disposition de l'apprenant différentes procédures, options, et commandes différentes lui permettant d'atteindre un même objectif.

Cependant, une interface ne peut convenir à la fois à tous ses utilisateurs potentiels. Pour éviter les effets négatifs sur les apprenants, une interface doit s'adapter à chaque apprenant d'entre eux (Bastien et Scapin, 1997).

4.2.4.1 Flexibilité

Elle se réfère aux moyens offerts à la disposition de l'apprenant pour personnaliser son interface et à la capacité de cette dernière à s'adapter au problème à résoudre. De cette

manière, les stratégies de travail et/ou les habitudes et les exigences des tâches sont prises en compte. Dans le tableau ci-après sont groupées les questions permettant de déterminer si ce sous-critère est satisfait par le STI (Bastien et Scapin, 1997).

Check-list	X	0	1	2	Commentaires
E.4.1.1 Les couleurs de l'interface sont modifiables de façon à s'adapter au goût et habitudes de l'apprenant					
E.4.1.2 L'interface s'adapte au contexte du problème à résoudre					
E.4.1.3 Il y a des raccourcis permettant d'éviter l'entrée de longues commandes					
E.4.1.4 Les options de menu sont dépendantes du contexte					
E.4.1.5 Un guidage flexible de l'apprenant est fourni					
E.4.1.6 Il permet de zoomer l'affichage					
E.4.1.7 L'apprenant peut nommer les commandes					
E.4.1.8 L'apprenant peut personnaliser l'affichage des fenêtres					
E.4.1.9 La séquence de contrôle est flexible					
E.4.1.10 La conception de la saisie des données est flexible					

Tableau 4.2.4.1.1 Évaluation de l'ergonomie : la flexibilité

4.2.4.2 Prise en compte de l'expérience

Ce critère concerne l'ensemble des moyens disponibles pour prendre en compte le niveau d'expérience de l'apprenant (expérimenté, débutant, occasionnel). L'expérience dont il est question ici est le niveau de familiarité de l'apprenant par rapport à la manipulation du système. Le tableau ci-après présente les questions auxquelles doit répondre un STI pour satisfaire ce sous-critère.

Check-list	X	0	1	2	Commentaires
E.4.2.1 Le système tient compte de la familiarité, de l'expérience de l'apprenant					
E.4.2.2 Le niveau de guidance du système diminue avec l'expérience de l'apprenant					
E.4.2.3 Le niveau de contrôle est adapté au niveau de l'apprenant (débutant à expert)					

Tableau 4.2.4.2.1 Évaluation de l'ergonomie: La protection contre les erreurs

4.2.5 Gestion des erreurs

Le critère Gestion des erreurs concerne l'ensemble des moyens permettant l'évitement ou la réduction des erreurs, et leur correction lorsqu'elles surviennent. Les erreurs dont il est question ici sont les erreurs de manipulation du système telles que : les saisies de données incorrectes, les saisies dans des formats inadéquats, les saisies de commandes avec une syntaxe incorrecte, etc.... Ces erreurs nuisent au bon fonctionnement du système. Il ne faut pas confondre avec les erreurs de résolution d'un problème posé par le STI, car ces dernières sont normales et seront traitées en fonction de la stratégie pédagogique employée par le STI.

Les interruptions provoquées par les erreurs manipulations de l'apprenant nuisent au bon déroulement des activités des apprenants. De façon générale, elles causent un allongement des transactions et perturbent la planification. Ainsi, la réduction des erreurs améliorerait la performance dans la mesure où elle permettrait une réduction des interruptions en raison de cela, le système doit protéger l'apprenant contre ce genre d'erreurs, l'en avertir et permettre leur correction.

4.2.5.1 La protection contre les erreurs

Elle s'intéresse aux moyens disponibles permettant la prévention et la détection des erreurs d'entrée de données, les erreurs de commandes, ou les actions à conséquences destructrices. La détection des erreurs doit être effectuée de préférence avant la validation, car cela perturbe moins le déroulement de la transaction. Dans le tableau ci-après nous présentons les différentes questions auxquelles devra répondre un STI satisfaisant ce sous-critère :

Check-list	X	0	1	2	Commentaires
E.5.1.1 Si la base de données inclut un groupe de données, l'apprenant peut saisir plus qu'un groupe dans le même écran					
E.5.1.2 Les points ou les soulignements sont utilisés pour indiquer la longueur des champs					
E.5.1.3 Le nom du choix de menu de haut niveau est utilisé comme titre dans les menus de bas niveaux					
E.5.1.4 Le choix des menus est logique, distinct et mutuellement exclusif					
E.5.1.5 Les champs de saisie de données sont vides lorsque possible					
E.5.1.6 Lorsque le système affiche plusieurs fenêtres, ces dernières sont visibles et il est simple de naviguer entre elles					
E.5.1.7 Les boutons/touches de fonctions qui peuvent causer des conséquences sérieuses sont difficilement atteignables					
E.5.1.8 Les boutons/touches de fonctions aux conséquences les plus sérieuses sont éloignés de ceux/celles aux conséquences moins sérieuses ou de ceux/celles fréquemment utilisés					
E.5.1.9 Le système prévient l'apprenant des erreurs que ce dernier peut ou va commettre lorsque possible					
E.5.1.10 Le système interprète intelligemment la variation de commande de l'apprenant					

Tableau 4.2.5.1.1 Évaluation de l'ergonomie: La protection contre les erreurs

4.2.5.2 La qualité des messages d'erreur

Elle s'intéresse à l'expression et au contenu des messages d'erreurs. Il s'agit plus précisément de s'assurer que les messages d'erreurs sont : pertinents, facilement lisible, précis quant à la nature des erreurs et aux actions correctives. Ainsi, les questions auxquelles doit répondre un système pour satisfaire ce sous-critère sont présentées dans le tableau ci-après :

Check-list	X	0	1	2	Commentaires
E.5.2.1 Les messages d'erreurs sont lisibles					
E.5.2.2 Concernant la nature de l'erreur commise, les messages d'erreurs sont précis					
E.5.2.3 Les messages sont précis en ce qui concerne les actions pour les corriger					
E.5.2.4 Le système utilise des effets visuels appropriés pour signaler une erreur dans la saisie d'un champ de données					
E.5.2.5 Le système utilise des effets sonores appropriés pour signaler une erreur dans la saisie d'un champ de données					
E.5.2.6 Les messages d'erreurs évitent l'utilisation de mots violents ou hostiles envers l'utilisateur					
E.5.2.7 Les messages d'erreurs sont formulés de façon à blâmer le système et non l'apprenant					
E.5.2.8 Le système informe l'apprenant sur la gravité de l'erreur commise					
E.5.2.9 La structure des menus matche avec la structure des tâches					

Tableau 4.2.5.2.1 Évaluation de l'ergonomie: La qualité des messages d'erreurs

4.2.5.3 La correction des erreurs

Il s'agit de l'ensemble des moyens offerts à l'utilisateur pour corriger immédiatement ses erreurs. Nous rappelons que les erreurs dont nous parlons ici sont des erreurs de saisie, de

manipulation et non pas des erreurs de résolution du problème. En effet, les erreurs liées à la résolution de problème sont liées à la stratégie pédagogique du système et sont donc indépendantes de son ergonomie. Le tableau ci-après présente les exigences auxquelles doit répondre un système pour satisfaire ce sous-critère ergonomique

Check-list	X	0	1	2	<u>Commentaires</u>
E.5.3.1 Le STI permet la correction des erreurs de saisie ou de manipulation					
E.5.3.21 Le STI a un système de traçage qui permet à l'apprenant de faire plusieurs retours arrière (« undo ») de saisie ou manipulation					
E.5.3.3 Le STI a un système de traçage permettant de faire revenir des saisies ou manipulations annulées (« redo »)					

Tableau 4.2.5.3.1 Évaluation de l'ergonomie: La correction des erreurs

4.2.6 Homogénéité/cohérence

Ce critère s'intéresse à la manière avec laquelle les choix de conception de l'interface (codes, dénominations, formats, procédures, etc.) sont maintenus pour des contextes similaires, et sont différents pour des contextes différents.

En effet, la stabilité de la localisation, d'utilisation, ainsi que du format, ou de la syntaxe des procédures, étiquettes, commandes, etc., d'un écran à l'autre, d'une session à l'autre facilite reconnaissances de ces derniers. Cela constitue un facteur de réduction des erreurs, car le système est plus prévisible et les apprentissages (de manipulation du système) plus généralisables. De plus, Il est notable que le manque d'homogénéité puisse causer une augmentation considérable du temps de recherche. En raison de cela, le manque d'homogénéité est aussi une raison importante de refus d'utilisation (Bastien et Scapin, 1997).

Un système satisfaisant ce critère répondra positivement aux questions du tableau ci-après

Check-list	X	0	1	2	Commentaires
E.6.1 Le format des procédures, des commandes est stable d'un écran à un autre, d'une session à une autre					
E.6.2 Les standards industriels ou les normes ont été bien respectés à travers tous les écrans du système					
E.6.3 Toutes les icônes sont étiquetées					
E.6.4 Il y a des effets visuels saillants permettant d'identifier les fenêtres actives					
E.6.5 Toutes les fenêtres ont un titre					
E.6.6 S'il y a « quitter » dans le choix de menu, il apparaît toujours en fin de liste					
E.6.7 Les instructions en ligne apparaissent à des endroits bien choisis dans tous les écrans du système					

Tableau 4.2.6.1 Évaluation de l'ergonomie: Homogénéité et cohérence

4.2.7 Signifiante des codes et dénominations

La signifiante des codes et dénominations concerne l'adéquation de la relation entre l'objet ou l'information affichée ou entrée, et son référent. Des codes et dénominations sont " signifiants " dans la mesure où ils bénéficient d'une relation sémantique forte avec leur référent.

En effet lorsque le codage et les dénominations sont significatifs, cela contribue à l'amélioration du rappel et de la reconnaissance diminuant ainsi le risque d'erreur. Dans le cas contraire, ils sont une source d'erreur pour les apprenants dans la mesure où ils peuvent

leur suggérer des opérations inappropriées et ainsi conduire à des erreurs (Bastien et Scapin, 1997). Le tableau ci-après liste les exigences auxquelles doit répondre un STI pour satisfaire à ce critère :

Check-list	X	0	1	2	Commentaires
E.7.1.1 La codification et la dénomination respecte celle du domaine enseigné					
E.7.1.2 Les dénominations et la codification des items sont significatives des actions qui en découlent					

Tableau 4.2.7.1 Évaluation de l'ergonomie: Signifiante des codes et dénomination

4.2.8 Compatibilité

La compatibilité concerne l'accord entre les caractéristiques des apprenants (mémoire, perceptions, habitudes, compétences, âge, attentes, etc.) et les tâches, d'une part, et l'organisation des sorties, des entrées et du dialogue d'une application donnée, d'autre part. De plus, ce critère s'intéresse également à la cohérence entre divers environnements ou applications (Bastien et Scapin, 1997).

En effet, la vitesse de transfert d'information intercontexte est de d'autant plus grande que la taille d'information qui doit être recoder par l'apprenant est réduite. Par ailleurs la présentation de l'information dans un format utilisable et familier à l'apprenant réduit l'effort d'interprétations ce qui améliore la performance. Les performances sont meilleures lorsque l'information est présentée sous une forme directement utilisable. Le tableau ci-après présente les exigences à satisfaire pour répondre à ce critère :

Check-list	X	0	1	2	Commentaires
E.8.1 Le STI présente des similitudes d'interface avec d'autres STI du domaine					

E.8.2 Les mots utilisés par le STI sont familiers au domaine					
E.8.3 L'organisation des sorties est conforme au champ d'application du système					

Tableau 4.2.8.1 Évaluation de l'ergonomie : compatibilité

4.3 Grille d'évaluation de l'utilité d'un système tutoriel intelligent

Pour l'évaluation de l'utilité, nous avons fait ressortir les critères les plus importants que devait avoir chaque modèle (modèle apprenant, modèle pédagogique, modèle base de connaissance et modèle expert).

4.3.1 Le modèle apprenant

Comme nous l'avons précédemment mentionné, le modèle de l'apprenant sert à diagnostiquer et modéliser la compréhension de l'apprenant (voir section 3.4.1.3). Ce dernier peut être visible ou non par l'apprenant et calibrable ou non. Ainsi, le modèle apprenant d'un STI peut être comparé à l'apprenant lui-même (dont les connaissances dans le temps peuvent être changeantes, contradictoires et difficiles à observer). Conséquemment, les critères déterminants de l'évaluation d'un modèle apprenant sont la validité, la fiabilité. Le critère de fiabilité n'étant évaluable que par une méthode expérimentale, nous nous limiterons à celui de la validité, car notre grille se veut être une grille d'inspection.

Le modèle de l'apprenant est valide dans la mesure où il reflète correctement l'apprenant dans le temps. Cet objectif n'est réalisable pleinement que dans la mesure où le système tient compte de tous les facteurs pouvant influencer le raisonnement de l'apprenant. Dans le tableau ci-après sont regroupées les questions auxquelles devrait satisfaire le modèle apprenant d'un bon STI.

Check-list	X	0	1	2	Commentaires
U.1.1 Le modèle apprenant est explicitement défini dans le système					
U.1.2 Le modèle apprenant est visible par l'apprenant lui-même					
U.1.3 Le modèle apprenant peut se calibrer automatiquement					
U.1.4 L'apprenant peut calibrer personnellement le modèle apprenant					
U.1.5 Le calibrage permet de se faire une idée de(s) apprenants(s) dès le départ					
U.1.6 L'information diagnostiquée obtenue par le système sur les connaissances de l'apprenant est identique aux diagnostics obtenus indépendamment du système à tout moment					
U.1.7 Le système tient compte des erreurs récurrentes de/ des apprenants (modèle cognitif: overlay par perturbation)					
U.1.8 Le système tient compte des informations affectives pouvant influencer le comportement de/des apprenants (modèle affectif de l'apprenant)					
U.1.9 Le système tient compte du contexte d'apprentissage					
U.1.10 Le système fait un traçage du raisonnement de/des apprenants (« modèle de tracking »)					
U.1.11 Le système analyse les événements d'apprentissage pour déterminer les connaissances qui y sont employées (« knowledge tracking »)					

U.1.12 Les apprenants peuvent collaborer entre eux pour résoudre un problème (approche collaborative)					
---	--	--	--	--	--

Tableau 4.3.1 Évaluation de l'utilité : le modèle apprenant

4.3.2 Le modèle pédagogique

Comme nous l'avons vu dans le chapitre précédent, un objectif des développeurs des STI est de construire un système qui crée un enseignement personnalisé en accord avec les caractéristiques des apprenants (leurs connaissances de base, leur niveau de développement cognitif, leurs styles d'apprentissage), dans le but d'améliorer l'efficacité de l'apprentissage de l'apprenant. Les standards auxquels le modèle pédagogique peut être comparé sont les théories d'enseignement et les tuteurs humains. Or, les théories de l'enseignement et les méthodes d'enseignants humains changent avec le temps et le contexte où elles s'appliquent. Tricot a retenu des critères d'évaluation de l'utilité dont certains s'appliquent très bien à l'évaluation du modèle pédagogique à savoir : la précision/présentation des objectifs pédagogiques et les adéquations des scénarios pédagogiques.

4.3.2.1 Précision/présentation des objectifs et identification des techniques pédagogiques employées

Les objectifs pédagogiques doivent être précisés, afin que l'apprenant sache à quoi s'attendre du système. De plus, il est important d'identifier les techniques pédagogiques utilisées par le système car c'est à partir d'elles que l'adéquation des scénarios didactiques est jugée. Dans le tableau ci-après sont regroupées les questions permettant de savoir si les objectifs didactiques du système sont présents et surtout de déterminer les approches didactiques utilisées par le STI.

Check-list	X	0	1	2	Commentaires
U.2.1.1 Le système présente de façon précise ses objectifs pédagogiques					

U.2.1.2 Le système intervient lorsque l'apprenant s'éloigne de la solution (coaching)					
U.2.1.3 Le système interroge l'apprenant sur ses erreurs afin de le faire progresser (approche socratique)					
U.2.1.4 Le système essaie d'induire l'apprenant en erreur (approche par perturbation)					
U.2.1.5 Le système permet à l'apprenant d'apprendre en modifiant des paramètres et en observant les conséquences de ses actions dans un environnement simulé (guidage)					
U.2.1.6 Le ou les approches pédagogiques utilisées par le système sont pertinentes par rapport au domaine enseigné					
U.2.1.7 Le tuteur peut changer d'humeur en fonction de la qualité du raisonnement de l'apprenant (réalité virtuelle)					
U.2.1.8 Si le système a plusieurs approches pédagogiques, l'alternance des approches est pertinente					

Tableau 4.3.2.1.1 Évaluation de l'utilité : le modèle pédagogique- approche et objectifs

4.3.2.2 Adéquation des scénarios didactiques

Les objectifs pédagogiques et les scénarios ne doivent pas être discordants, mais aller de pair. De plus les connaissances de l'apprenant étant mouvantes, il est important que les scénarios didactiques en tiennent compte pour permettre une évaluation adéquate de l'apprenant. Dans le tableau ci-après nous avons regroupé les questions permettant de déterminer si les scénarios didactiques sont adéquats ou non.

Check-list	X	0	1	2	Commentaires
U.2.2.1 Les scénarios didactiques sont conformes aux objectifs pédagogiques du système					
U.2.2.2 Les scénarios didactiques tiennent compte de l'évolution des connaissances/habiletés de l'apprenant					
U.2.2.3 Les scénarios didactiques tiennent compte de la régression des connaissances/habiletés de l'apprenant					
U.2.2.4 Les scénarios didactiques tiennent compte du contexte d'apprentissage					
U.2.2.5 Les diagnostics du modèle pédagogique sont réalistes					

Tableau 4.3.2.1.1 Évaluation de l'utilité : le modèle pédagogique- adéquations des scénarios didactiques

4.3.3 Le modèle du domaine ou Base de connaissance

Ce modèle contient les connaissances et/ou habiletés que le STI souhaite faire acquérir à l'apprenant. Idéalement, la précision de cette composante doit être vérifiée avant que le système ne soit complété et assumée après. Pour évaluer ce modèle, le critère de conformité au domaine mentionné par Marc et Greer (voir section 3.4.1.1) est des plus déterminants. En effet, on ne saurait prétendre utiliser le langage naturel pour faciliter l'apprentissage si on n'est pas capable de l'analyser. Ainsi pour l'évaluation de cette composante, l'important est de déterminer si le type de représentation de connaissances permet la réalisation des objectifs pédagogiques du système. En d'autres termes, pour évaluer ce modèle il est important d'identifier le(s) type(s) de représentation des connaissances utilisé(s) et de s'assurer de sa (leur) pertinence par rapport aux objectifs didactiques.

Par ailleurs, il est important de noter que nous sommes de nos jours dans une ère de communication et d'échanges. Une base de connaissance moderne doit par conséquent permettre d'acquérir de nouvelles connaissances voire de partager ces connaissances avec d'autres bases si possible. Il est par conséquent important de nos jours de faire preuve d'évolutivité et d'ouverture. Dans le tableau ci-après nous sont groupées les questions permettant de juger de la pertinence de la base de connaissance du STI évalué.

Check-list	X	0	1	2	Commentaires
U.3.1 Le système utilise des connaissances procédurales (si... alors...sinon)					
U.3.2 Le système utilise des connaissances déclaratives (faits, concept, évènement, images, proposition,...)					
U.3.3 Le système utilise des connaissances épisodiques					
U.3.4 Les connaissances sont retranscrites à l'aide d'un formalisme logique (règle de production, logique floue,...)					
U.3.5 Les connaissances sont retranscrites à l'aide d'un formalisme sémantique (réseaux sémantiques)					
U.3.6 Les connaissances sont retranscrites à l'aide d'un formalisme hybride (réseaux bayésiens)					
U.3.7 Les types de connaissances et les formalismes utilisés sont en accord avec les scénarios didactiques					
U.3.8 Le corpus du modèle de connaissance peut évoluer					
U.3.9 On a accès aux connaissances du système					

U.3.10 Le système peut partager ses connaissances avec d'autres systèmes					
--	--	--	--	--	--

Tableau 4.3.3.1 Évaluation de l'utilité : base de connaissance-conformité au domaine et facilitation de l'atteinte des objectifs pédagogiques

4.3.4 La modèle expert

Grâce à ce modèle, le système peut résoudre les problèmes soumis à l'apprenant. Le critère le plus déterminant pour l'évaluation de cette composante est la rectitude. En effet, cette composante doit être capable de résoudre avec exactitude les problèmes proposés. De plus, elle doit permettre de déterminer la cause exacte des erreurs de l'apprenant et si possible de les prédire. De ce fait, un bon modèle expert doit être capable de résoudre le problème et d'analyser avec exactitude les erreurs de l'apprenant.

Par ailleurs, plus un expert est proche du modèle cognitif, plus son raisonnement est exact. Dans le tableau ci-après sont groupées les questions permettant de s'assurer de la satisfaction de ce critère par un STI.

Check-list	X	0	1	2	<u>Commentaires</u>
U.4.1 Le système propose des solutions exactes aux problèmes proposés					
U.4.2 Le système détermine de façon précise le(s) cause(s) de(s) erreur(s) de l'apprenant					
U.4.3 Le système est capable de prédire les erreurs de l'apprenant à partir d'un début de raisonnement de l'apprenant					
U.4.4 On a accès aux connaissances du système expert (l'expert n'est pas une boîte noire)					

U.4.5 Les connaissances de l'apprenant sont comparées à celle du système expert (boîte de verre)					
U.4.6 Ce modèle est basé sur des théories de cognition pertinentes					

Tableau 4.3.4.1 Évaluation de l'utilité : modèle expert-rectitude

CHAPITRE V

PRÉSENTATION D'ANDES

ANDES est un système tutoriel intelligent mature aidant à l'apprentissage universitaire de la physique. Il remplace le crayon et le papier dans la résolution des devoirs (exercices à faire à la maison). Les cinq années d'évaluation de ce dernier à l'académie navale universitaire des États-Unis indiquent qu'il a amélioré l'apprentissage des apprenants. Ces cinq années d'évaluation d'ANDES lui ont permis d'atteindre un niveau de maturité qui n'a rien à envier aux autres STI. De plus ANDES est gratuitement accessible via l'Internet. Il est donc normal que nous voulions tester notre grille sur un STI de cette envergure qui est une référence du domaine.

ANDES est rendu présentement à sa deuxième version qui se distingue de la précédente par le fait qu'elle n'utilise plus les réseaux bayésiens. Les auteurs ont constaté qu'étant donné qu'ANDES2 est utilisé dans le cadre d'un cours, c'est l'enseignant qui décidera du niveau des exercices présentés à l'apprenant (VanLehn et al., 2005a). Dans le présent chapitre, nous allons présenter l'architecture d'ANDES2 et un résumé des évaluations qu'a subies ce dernier.

5.1 Architecture

Afin de présenter l'architecture de ANDES2, nous avons fait un résumé de l'article de VanLehn et al. (2005a) qui rapporte des informations détaillées sur l'architecture d'ANDES2.

5.1.1 Interface

5.1.1 Interface

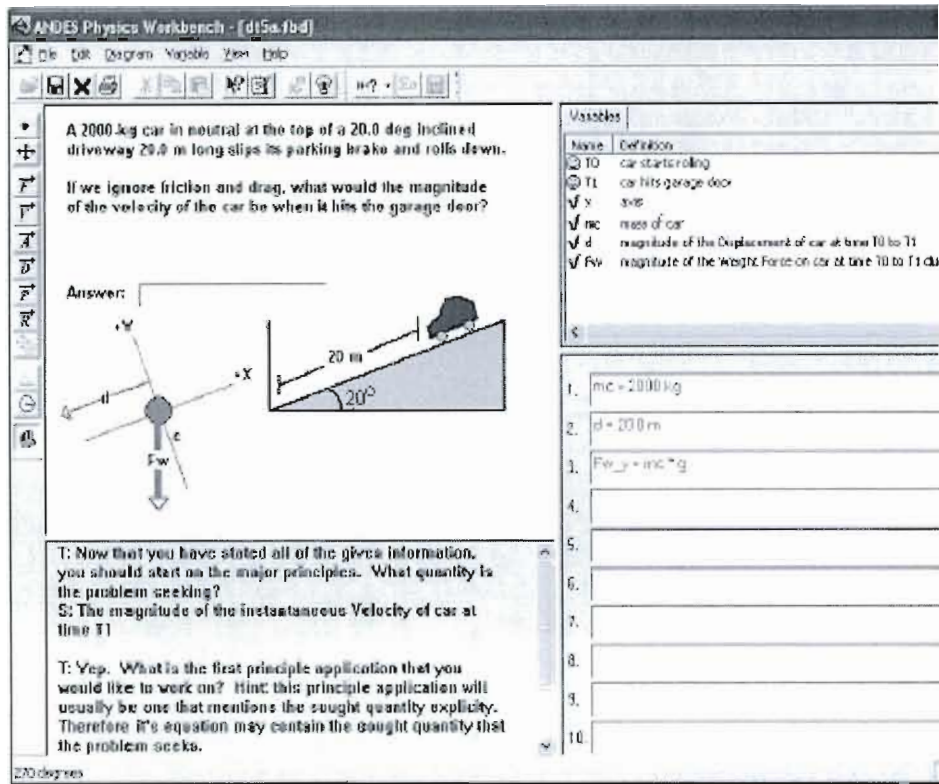


Figure 5.1.1.1 Interface ANDES2 (coupée à droite, (VanLehn *et al.*, 2005a))

Dans l'objectif de faciliter l'apprentissage, ANDES2 présente une interface qui veut se rapprocher au maximum de celle de la feuille de papier et du crayon de l'apprenant. La précédente figure est un exemple typique de problème de physique et de sa résolution avec ANDES2. L'apprenant lit le problème (dans le coin haut gauche de la fenêtre), dessine le vecteur de coordonnées (dans le coin bas gauche de la fenêtre), définit les variables (dans le coin haut droit de la fenêtre) et entre les équations (dans le coin bas droit de la fenêtre). L'apprenant exécute ainsi les actions similaires à celles effectuées sur une feuille de papier.

Toutefois, contrairement au papier, ANDES2 donne un feed-back aussitôt qu'une action est complétée. Les entrées sont colorées en vert si elles sont correctes et en rouge sinon. Ainsi,

dans l'exemple précédent (voir la figure 5.1.1.1) toutes les entrées sont correctes, donc en vert, à l'exception de la troisième entrée ($Fw_y=mc*g$), qui par conséquent est rouge.

En outre, il oblige l'apprenant à définir avant usage les vecteurs adéquats. Pour ce faire, les vecteurs et autres objets graphiques sont dessinés à l'aide de la souris, de la barre d'outils présente dans le coin gauche de l'interface (voir figure 5.1.1.1) et d'une boîte de dialogue (boîte qui sert à la définition de l'objet dessiné).

ANDES2 fournit un solveur d'équations. Ainsi, lorsque l'apprenant clique sur le bouton étiqueté « X=? », il demande quelle variable l'apprenant souhaite résoudre. À la suite de cela, il essaye de résoudre le système d'équations fournies par l'apprenant. S'il parvient à le résoudre, il affiche une équation de la forme $\langle \text{variable} \rangle = \langle \text{valeur} \rangle$. Ce solveur d'équations d'ANDES2 permet d'éviter les erreurs de manipulation d'une calculatrice (VanLehn *et al.*, 2005a).

Au niveau de l'aide, ANDES2 fournit trois types d'aide pour encadrer les apprenants dans leur apprentissage. Ainsi, les erreurs d'inattention sont signalées par des messages d'aide non sollicités (popup). Concernant les erreurs de compréhension, ANDES2 fournit de l'aide sur demande⁸ soit du genre « What's wrong help » qui permet d'obtenir des explications sur le problème concernant les entités marquées en rouge et d'un autre côté ou du genre « Next step help » qui permet à l'apprenant de demander au tuteur de le renseigner sur la prochaine étape à suivre.

A remarquer, ANDES2 se rapproche des WBH (web-based homework) standard. En effet, les apprenants peuvent soumettre leur solution via l'internet à leurs enseignants. Ces derniers peuvent accepter le résultat de l'évaluation d'ANDES2 ou faire leur propre évaluation.

⁸ Fournir de l'aide à la demande pour des problèmes de compréhension a pour but d'accroître les chances d'autocorrection des étudiants qui est un facteur consolidant de l'apprentissage d'après les théories constructivistes de l'apprentissage (e.g., Merrill, Reiser, Ranney, & Trafton, 1992).

5.1.2 Modèle pédagogique

Le modèle pédagogique consiste en l'implémentation de trois types d'aide fournie par ANDES2 à savoir : le feed-back immédiat, le « what's wrong help » et le « Next step help » ou « quelle est la prochaine étape? ». Le « What's wrong help » et le « Next step help » génèrent une séquence de conseils constituée de trois types qui sont, dans l'ordre :

- **l'indice** : Son but est de permettre à l'apprenant de localiser son erreur. De ce fait, l'autocorrection et l'apprentissage s'en trouvent facilités (Hume *et al.*, 1996 Merrill *et al.*, 1992). Ainsi ANDES2 peut par exemple suggérer à l'apprenant de vérifier sa trigonométrie.
- **le rappel de notions (« teaching hint »)**: Si l'indice s'avère insuffisant pour l'aider à résoudre son problème et que l'apprenant demande plus d'aide, ANDES2 pourra notamment décrire la formule à solutionner. Et pour éviter à l'apprenant d'avoir à lire de longs textes, tout rappel d'enseignement sera maintenu au plus bref possible.
- **le « bottom out hint »**: Ce dernier survient lorsque le type de conseil précédent a été insuffisant à éclairer l'apprenant et qu'il demande encore plus d'aide. Dans ce cas, ANDES2 va donner des instructions précises et brèves sur la façon exacte dont il faut s'y prendre pour résoudre le problème.

Il peut arriver qu'ANDES2 soit incapable d'inférer sur ce que l'apprenant souhaiterait faire. Dans ce cas, il va interroger l'apprenant avant de lui fournir de l'aide. La figure 5.1.1.1 en est une illustration. L'apprenant demande avec le « Next step help » la prochaine étape et ANDES2 lui pose la question : « Quelle est la quantité cherchée dans le problème? ». Une boîte de menu apparaît pour que l'apprenant fournisse la réponse. L'écho de la réponse de l'apprenant apparaît dans la fenêtre en bas à gauche.

Dans les sous-sections suivantes, nous décrirons de façon détaillée comment ces trois types d'aide fournie par ANDES2 sont implantés.

5.1.2.1 Implémentation du feedback immédiat

Le domaine d'ANDES2 étant la physique, l'apprenant peut par conséquent fournir deux types de données à savoir : les équations et les non-équations. Le dessin des vecteurs, des axes de coordonnées et les données scalaires sont considérés comme n'étant pas des équations. L'implémentation du feed-back immédiat tient compte de ces deux types de données que peut lui fournir l'apprenant.

Pour les non-équations, il va premièrement valider leur syntaxe (p. ex., il s'assure que le formulaire permettant de dessiner un axe est rempli correctement), puis il va vérifier que cette non-équation est une composante d'au moins une solution du problème.

Concernant les équations, comme dans le cas précédent ANDES2 va tout d'abord s'assurer de la validité de leur syntaxe. Cette validation est faite à l'aide d'un analyseur syntaxique indépendant du contexte. Ce processus de validation permet aussi la détection des variables non définies. Après la validation de la syntaxe, la dimension de l'équation est vérifiée. Cela permet d'éviter le cas où l'apprenant oublie d'indiquer l'unité d'un nombre dimensionnel. Par la suite, lorsque les équations des apprenants sont dimensionnellement consistantes, ANDES2 complète sa validation en utilisant une technique appelée couleur par nombre (« color by number », voir Shapiro, 2005).

Toutefois, il faut remarquer que l'algorithme de couleur par nombre est efficace mais pas parfait. Un problème connu de cette technique est que : une sous expression incorrecte dans l'équation d'un apprenant peut être ignorée si elle est multipliée par une expression de valeur zéro. Ce cas défectueux de la technique arrive rarement dans ANDES2 comme le démontre Shapiro vu que sur 5766 solutions de problèmes, il y a eu un seul cas où l'équation solution de l'apprenant était verte alors qu'elle était fausse (Shapiro, 2005).

5.1.2.2 Implementation du what's wrong help

Lorsqu'une donnée rouge est sélectionnée, l'apprenant peut cliquer sur le bouton « *what's wrong help* » (quel est le problème?) et ANDES2 va appliquer un grand ensemble gestionnaire d'erreur. Chaque gestionnaire d'erreur prend la donnée de l'apprenant en

paramètre. Si l'erreur est reconnue, il en retourne un conseil et une priorité. Dans le cas contraire, le gestionnaire d'erreur retourne nul. Si l'erreur est reconnue par plusieurs gestionnaires, ANDES2 choisit la séquence de conseils la plus prioritaire.

Le gestionnaire d'erreurs d'ANDES2 fonctionne en effectuant des modifications sur les données entrées par l'apprenant. Si les modifications génèrent une entrée correcte alors, le gestionnaire d'erreur construit une séquence de conseils se référant à la donnée saisie par l'apprenant. Dans le cas particulier des équations, le gestionnaire d'erreurs effectue les modifications et appelle l'algorithme de couleur par nombre sur l'équation modifiée. Comme illustration, supposons que pour résoudre le problème de la figure 5.1.1.1 l'apprenant saisisse l'équation $Fw_x = Fw \cdot \cos(20\text{deg})$. Le gestionnaire d'erreur va noter qu'il y a un cosinus dans l'équation et changer l'équation pour $Fw_x = Fw \cdot \sin(20\text{deg})$ et soumettre la nouvelle équation à l'algorithme de couleur par nombre qui va dire que l'équation est correcte. Conséquence, le gestionnaire d'erreur va générer la séquence de conseils présentée dans la figure ci-après :

- Check your trigonometry.
- If you are trying to calculate the component of a vector along an axis, here is a general formula that will always work: Let θ_V be the angle as you move counterclockwise from the horizontal to the vector. Let θ_x be the rotation of the x axis from the horizontal. (θ_V and θ_x appear in the Variables window.) Then: $V_x = V \cdot \cos(\theta_V - \theta_x)$ and $V_y = V \cdot \sin(\theta_V - \theta_x)$.
- Replace $\cos(20 \text{ deg})$ with $\sin(20 \text{ deg})$.

Figure 5.1.2.2.1 Exemple de séquence de conseil d'ANDES2 (VanLehn *et al.*, 2005a)

Il faut remarquer qu'ANDES2 ne teste pas les erreurs multiples sur une donnée entrée par l'apprenant. Ainsi, si nous reprenons notre précédent exemple: $Fw_x = -Fw \cdot \cos(20\text{deg})$. ANDES2 est incapable trouver qu'il y a une erreur de signe et une erreur de trigonométrie bien qu'il puisse trouver chacune de ces erreurs individuellement.

5.1.2.3 Implementation du next step help

Chaque problème dans ANDES2 est associé à un de graphe de solution. Ce dernier est une structure de données qui représente le plan hiérarchique de chacune des solutions possibles du problème à résoudre avec les interconnexions qu'elles ont entre elles. Cette structure de données est un graphe orienté qui est le graphe des solutions du problème. La figure 5.1.2.2.1 est un exemple partiel du graphe de l'information du graphe de solution du problème de la figure 5.1.1.1. Cependant, cette figure ne présente pas le fait qu'une branche part de l'étape 3 et se rend à la fois à l'étape 6 et à l'étape 4. Cela signifie qu'il y a des solutions alternatives. Ces branches convergent plus tard; les étapes 5 et 6 pointent toutes les deux sur l'étape 7.

1. Draw the body, which defines the mass variable, m_c .
2. Define coordinate axes, rotated 20 degrees
3. Define given quantities
 - 3.1. Draw the displacement, d
 - 3.2. Enter the given value of displacement, $d = 20 \text{ m}$
 - 3.3. Enter the given value of mass, $m_c = 2000 \text{ kg}$
4. Apply translational kinematics and get $v_f^2 = 2 \cdot a \cdot d$
 - 4.1. Draw the vector diagram
 - 4.1.1. Draw the initial velocity, v_i
 - 4.1.2. Draw the final velocity, v_f
 - 4.1.3. Draw the acceleration, a
 - 4.1.4. Draw the displacement, d (shared with step 3.1)
 - 4.2. Apply the fundamental principle: $v_f^2 = v_i^2 + 2 \cdot a \cdot d$
 - 4.3. Project the vectors onto the x-axis
 - 4.3.1. Apply projection: $v_{i,x} = 0$
 - 4.3.2. Apply projection: $v_{f,x} = -v_f$
 - 4.3.3. Apply projection: $a_x = -a$
 - 4.3.4. Apply projection: $d_x = -d$
5. Apply Newton's second law and get $a = -g \cdot \cos(200 \text{ deg})$
 - 5.1. Draw a free-body diagram
 - 5.1.1. Draw the force of gravity, F_w
 - 5.1.2. Draw the normal force, F_n
 - 5.1.3. Draw the acceleration, a (shared with step 4.1.3)
 - 5.2. Apply the fundamental principle: $m_c \cdot a_x = F_{w,x} + F_{n,x}$
 - 5.3. Project the vectors onto the x-axis
 - 5.3.1. Apply projection: $F_{n,x} = 0$
 - 5.3.2. Apply projection: $F_{w,x} = F_w \cdot \cos(250 \text{ deg})$
 - 5.3.3. Apply projection: $a_x = -a$
 - 5.4. Apply the weight law, $F_w = m_c \cdot g$
6. Apply conservation of energy and get $v_f^2 = 2 \cdot g \cdot d \cdot \sin(20 \text{ deg})$
 - 6.1. Define the kinetic energy at time T1: K_1
 - 6.2. Define the potential energy due to gravity at T0: P_0
 - 6.3. Apply the fundamental principle, $P_0 = K_1$
 - 6.4. Define height, h
 - 6.5. Apply the definition of gravitational potential energy: $P_0 = m_c \cdot g \cdot h$
 - 6.6. Apply trigonometry: $h = d \cdot \sin(20 \text{ deg})$
 - 6.7. Draw the final velocity, v_f (shared with 4.1.2)
 - 6.8. Apply the definition of kinetic energy, $K_1 = \frac{1}{2} \cdot m_c \cdot v_f^2$
7. Solve the system of equations for v_f and get $v_f = 11.59 \text{ m/s}$
8. Enter the value of v_f into the answer box

Figure 5.1.2.3.1 Graphe de solution du problème de la figure 5.1.1

De façon basique le « *next step help* » consiste à sélectionner la prochaine étape du graphe de solution pour fournir à l'apprenant la séquence de conseils associée à cette étape. Cette sélection d'étapes se base sur deux principes de conception. D'une part, le graphe est parcouru en profondeur et la première étape que l'apprenant n'a pas encore réalisée est sélectionnée. D'autre part, face à un embranchement dans le graphe de solutions, c'est la

branche du graphe de solutions ayant la plus grande proportion d'étapes effectuées par l'apprenant qui est sélectionnée.

Une fois l'étape à conseiller choisie, il génère la séquence de conseil composé de deux sous-séquences. La seconde sous-séquence est un conseil sur l'étape elle-même. Elle est constituée d'un rappel de notions et d'une explication de ce qu'il faut faire (« bottom out hint »). La première sous-séquence n'est générée que lorsqu'on est sur une étape principale. Elle s'assure que l'apprenant est au courant de l'étape principale. Cela peut impliquer qu'on peut soit : demander l'étape principale à l'apprenant (« quelle est la quantité cherchée par le problème »), dire en quoi consiste l'étape principale à l'apprenant (« Appliquons la seconde loi de Newton sur la voiture suivant l'axe des x), où se référer à un dialogue précédent (« Pourquoi ne continuez-vous pas d'appliquer la loi de conservation d'énergie à la voiture? »).

Pour réaliser sa politique, ANDES2 doit « mapper » les entrées de l'apprenant avec les étapes du graphe de solutions afin de marquer les étapes réalisées. Cette tâche est facile à réaliser dans le cas des entrées qui ne sont pas des équations, mais extrêmement complexe dans le cas contraire. Ainsi, si l'apprenant entre $Fw_x = -m \cdot g \cdot \sin(20 \text{ deg})$, alors ANDES2 doit trouver comment le « mapper » à l'étape 5.3.2 (writing $Fw_x = Fw \cdot \cos(250 \text{ deg})$) et à l'étape 5.4 (writing $Fw = m \cdot g$). ANDES1 dans sa précédente version avait tenté de résoudre ce problème en précompilant toutes les combinaisons algébriques possibles d'étapes. Cependant, cela devenait infaisable à mesure que les problèmes devenaient plus nombreux et surtout plus complexes.

Dans ANDES2, une solution plus adéquate a été trouvée (Shapiro, 2005). L'algorithme s'appelle « *indy check* » parce qu'il consiste à vérifier l'indépendance d'un ensemble multidimensionnel de vecteurs. Shapiro (2005) a montré que l'algorithme était juste sauf dans un cas spécial. Bien que discutablement rare, ce cas spécial est fréquent dans de cas

particuliers des problèmes utilisés par ANDES2. Une heuristique ne causant pas de perte visible de performance est utilisée pour détecter ce cas spécial⁹ (Shapiro, 2005).

En outre, une méthode de sélection d'étapes à conseiller devrait considérer deux critères : éviter de donner des conseils sur une étape déjà faite, et essayer de conseiller sur une étape pertinente par rapport à l'objectif de l'apprenant. De ce point de vue concernant l'identification des étapes déjà faites, ANDES2 est plus lent que les autres systèmes similaires. Cela s'explique par le fait que le graphe de solutions contient des primitives que l'apprenant peut entrer comme une combinaison. Ainsi, cette difficulté est la même pour tout domaine de tâches et interfaces usagers permettant une combinaison d'étapes primitives.

Toutefois, ANDES2 est plus efficace que les autres systèmes dans la résolution du problème de sélection de l'étape la plus appropriée aux objectifs courants de l'apprenant. Ce problème qui est d'ordre général est appelé problème de reconnaissance plan: on a un ensemble d'actions données, il faut trouver une sélection de plans hiérarchiques d'une librairie qui correspond le mieux à la séquence observée, en assumant que les plans sont en cours et que l'apprenant a toujours l'intention de réaliser ses objectifs. Ce problème reste encore un défi considéré comme insoluble (VanLehn, 1990).

Il est intéressant de noter qu'ANDES2 effectue une reconnaissance de plan limité dans deux cas. Ces derniers surviennent lorsque le problème a des solutions mutuellement exclusives. Ainsi, le premier cas survient si l'apprenant a commencé à entrer les étapes d'une ou

9

La version la plus commune du cas spécial apparaît quand le graphe de solution a des équations avec $\sin(\theta)$ ou $\cos(\theta)$, et il contient $\theta=n*90$ où n est un entier. Quand cela arrive, *indy check* est souvent incapable d'inclure $\theta=n*90^\circ$ comme une équation du graphe de solution qui caractérise l'équation de l'apprenant. Cela peut amener ANDES à penser que l'apprenant n'a pas encore entré $\theta=n*90^\circ$. Cependant, les apprenants n'écrivent pas présentement ce type d'équations. Les équations entrées ont pour effet secondaire de dessiner des vecteurs. Étant donné qu'ANDES2 sait exactement quel vecteur a été dessiné, il va éviter de choisir un vecteur qui a déjà été dessiné comme prochaine étape qu'importe ce que l'algorithme *indy check* dit. En principe d'autres versions du cas spécial peuvent survenir, mais aucune n'a encore échappé à l'heuristique et causé un échec du « *next step help* ».

plusieurs solutions. Dans ce cas, ANDES2 va suggérer les étapes de la solution où l'apprenant a le plus progressé qui serait équivalent à la solution la plus courte des solutions. Concernant le second cas, il survient lorsque l'apprenant n'a commencé aucune des solutions mutuellement exclusives. Pour ce cas, ANDES2 va demander à l'apprenant de spécifier le plan à suivre. En particulier, ANDES2 peut solliciter le choix de l'apprenant dans la première partie d'une séquence de conseil en demandant «Quel est la quantité recherchée par le problème? » et « Quel principe doit-on appliquer pour le trouver? ».

5.1.3 Base de connaissance et modèle expert

La base de connaissance d'ANDES2 est constituée de 356 problèmes et de 550 règles de physique. Pour chaque problème, il précompile le graphe de solution en utilisant un système expert. La littérature sur les différences entre l'expert et le novice dans la résolution de problèmes suggère que les experts organisent leurs solutions autour de principes majeurs. Ainsi, pour chacun de ces principes, la base de connaissance d'ANDES2 a une méthode de résolution de problème.

Une méthode de résolution de problème (MRP) incorpore un plan hiérarchique d'étapes permettant l'application d'un principe majeur combiné avec les principes mineurs qui lui sont associés. La figure ci-après illustre l'application de la MRP à la seconde loi de Newton. L'application de cette MRP au problème de la figure 5.1.1.1 produit l'étape principale 5 du graphe de solution précédemment décrit.

Par ailleurs, en plus des MRP, la base de connaissance d'ANDES2 contient plusieurs petites connaissances représentées comme de clauses de Horn¹⁰. Ces dernières permettent la gestion

¹⁰ Expression conditionnelle utilisée en programmation logique et comportant une seule conclusion. Les clauses de Horn sont ainsi nommées d'après le nom du logicien Alfred Horn qui, le premier, les a étudiées. Le langage Prolog est basé sur l'utilisation de clauses de Horn. Ces dernières se divisent en trois types : les faits, les contraintes d'intégrité entre les requêtes et les bases de données ou de connaissances dans les systèmes évolués.

des inférences telles que : “SI < objet 1> tire sur < objet 2> alors il y a une force de tension sur < objet 2> causée par < objet 1>.”

To apply Newton's second law to <body> at <time> along <axis>:

1. Draw a free-body diagram for <body> at <time> including <axis>
 - 1.1. For each force on <body> at <time>, draw it.
 - 1.2. Draw the acceleration of <body> at <time>
2. Write the fundamental equation in terms of components along <axis>
3. For each vector (i.e., acceleration and each force), write a projection equation for the vector along <axis>
4. For each minor principle of the form <force magnitude> = <expression> where the force is one of the ones on <body> at <time>, write the minor principle's equation.

Figure 5.1.3.1 Méthode de résolution de problème pour la seconde loi de NEWTON

La résolution d'un problème s'effectue à travers une forme de recherche spatiale d'état. Chaque état contient : un ensemble d'équations déjà générées, un ensemble de quantités qui sont « connues » et un ensemble de quantités « recherchées ». Lors de l'initialisation de ces états; il y a une équation pour chaque quantité donnée, les quantités données sont incluses dans l'ensemble des quantités connues et les quantités cherchées dans un problème incluent l'ensemble des quantités cherchées.

La recherche s'effectue en choisissant itérativement une MRP qui contient l'une des quantités recherchées et en appliquant la MRP. Quand la MRP a été appliquée et qu'un ensemble d'équations a été généré, l'algorithme indy check est appelé. Cela permet de s'assurer que de nouvelles équations ne puissent être générées. Ainsi, il est garanti que les équations dans les états sont indépendantes (on ne peut pas en générer une combinaison algébrique). Si l'indy check réussit, alors un nouvel état est créé en y ajoutant de nouvelles équations et en y effectuant la mise à jour des états connus et des états cherchés. La recherche continue jusqu'à ce que l'ensemble des quantités cherchées devienne vide. L'état final de chaque équation doit contenir N quantités connues et N équations indépendantes. Suite à cela, le résolveur d'équations peut être appelé pour produire un point de solution.

Cette recherche est effectuée de façon exhaustive. De cette manière, toutes les solutions possibles du problème sont générées. Dans le cas de la figure 5.1.1, cela résulterait en deux états finaux. Ces derniers sont fusionnés pour produire le graphe de solution précédemment présenté dans la figure 5.1.2.2.1

5.1.4 Modèle de l'apprenant

L'objectif du modèle apprenant est de fournir un diagnostic sur les compétences de l'apprenant. À partir de ce dernier, le STI peut offrir à l'apprenant une sélection de problèmes correspondant le mieux à ses besoins. Cela constitue ce qu'on appelle la macro-adaptation (Shute, 1993). Ainsi, grâce aux informations fournies par le modèle, le système peut décider de faire passer l'apprenant au chapitre suivant.

Le fait que ce soit le système qui décide de la progression de l'apprenant à travers les chapitres de la matière est très pertinent pour des cours destinés à des apprenants autodidactes. En effet, ces derniers contrairement aux apprenants « classiques » ne sont pas soumis aux contraintes horaires d'un cours classique, ils passent les examens uniquement quand ils se jugent prêts. Par conséquent, la gestion automatique du parcours de l'apprenant est idéale pour les apprenants autodidactes.

Si cette macro-adaptation est idéale pour les autodidactes, elle est très contraignante pour les apprenants « classiques ». Le problème ici vient du fait que contrairement aux autodidactes, ils disposent de moins de temps et se doivent de suivre le rythme de leur session académique. Ce problème a posé un frein à l'adoption répandue de l'autogestion du parcours de la matière pour les STI destiné aux apprenants « classiques » (Kulik et Bangert-Drowns, 1990).

ANDES2 n'étant pas encore utilisé dans des cours destinés aux apprenants autodidactes, il ne supporte pas la macro-adaptation. Par conséquent il n'a pas besoin d'un modèle bayésien d'ANDES1 qui devient inutile en raison du contexte. Toutefois, ANDES2 conserve les

évaluations des devoirs des apprenants. Ces évaluations ont pour unique but de permettre aux apprenants d'avoir une estimation de leur niveau.

5.2 Résumé de l'évaluation empirique

ANDES2 a été évalué dans le cadre du cours d'introduction à la physique de l'académie navale des Etats-Unis chaque session Automne de 1999 à 2003 (VanLehn et al., 2005). Dans cette section nous allons décrire les 5 évaluations et leurs résultats. Il est à noter que les informations présentées dans cette section proviennent de l'article de VanLehn et al. 2005b.

ANDES2 a été utilisé comme partie intégrante d'un cours académique normal de physique. Le cours avait de multiples groupes et chaque section du cours était dispensée par un enseignant différent. Les apprenants de tous les groupes avaient le même examen final et utilisaient le même manuel. Cependant, les différents enseignants donnaient des devoirs différents et différents examens d'une heure. Les examens d'une heure se faisaient en classe et mensuellement. Dans les groupes enseignés par les auteurs (Shelby, Treacy and Wintersgill) les apprenants étaient encouragés à faire leurs devoirs sur ANDES2. Chaque année, les groupes des collègues des auteurs étaient recrutés pour servir de contrôle.

Les groupes de contrôle effectuaient un devoir similaire, mais non identique à ceux effectués par les apprenants d'ANDES2. Des pratiques ont été mises au point et appliquées pour s'assurer que ces derniers prennent leurs devoirs au sérieux et les résolvent avec une bonne démarche (devoirs sont notés en se basant sur l'effort et la démarche, des jeux-questionnaires basés sur les devoirs précédents notés). Les apprenants ont tous eu le même examen final qui durait 3 heures et comptait une cinquantaine de choix multiples. Les examens mensuels quant à eux étaient constitués de 4 problèmes qu'il fallait résoudre en 1 heure. L'examen final était plus facile que ces derniers dans la mesure où il était fait de mini problèmes (4 minutes pour répondre à chaque question). De plus, l'examen d'une heure n'exigeait pas que les apprenants

détaillent l'intégralité de leur démarche. Nous discuterons en premier lieu de l'examen d'une heure (Vanlehn *et al.*, 2005b).

Year	1999	2000	2001	2002	2003	Overall
Andes students	173	140	129	93	93	455
Control students	162	135	44	53	44	276
Andes mean (SD)	73.7 (13.0)	70.0 (13.6)	71.8 (14.3)	68.2 (13.4)	71.5 (14.2)	0.22 (0.95)
Control mean (SD)	70.4 (15.6)	57.1 (19.0)	64.4 (13.1)	62.1 (13.7)	61.7 (16.3)	-0.37 (0.96)
P(Andes= Control)	0.036	<.0001	.003	0.005	0.0005	<.0001
Effect size	0.21	0.92	0.52	0.44	0.60	0.61

Tableau 5.2.1 Résultats pour l'examen d'une heure (Vanlehn *et al.*, 2005b).

Le tableau ci-dessus nous montre le résultat des 5 ans d'évaluation d'ANDES2. Il représente le taux de réussite moyen sur un ou plusieurs examens par année. Dans toutes les années, les apprenants d'ANDES2 ont eu une meilleure note que les apprenants de contrôle avec un effet de taille légèrement élevé (l'effet de taille est la moyenne d'ANDES2-la moyenne de contrôle/ l'écart type de la moyenne d'ANDES2 – la moyenne de contrôle.). L'évaluation de l'année 1999 a un effet de taille plus petit parce qu'à cette époque ANDES2 avait peu de problèmes de physique et plusieurs bogues. Cela était un facteur démotivant dans l'utilisation d'ANDES2. Cette année n'est pas par conséquent représentative de l'effet d'ANDES2.

Par ailleurs dans le but de calculer le résultat d'ensemble, les scores des examens ont été normalisés parce que les examens avaient de grandes moyennes différentes selon les années. Ainsi, toutes les notes des apprenants ont été converties en notes Z (note Z = (note – écart type) ÷ grand écart type). L'agrégation des notes Z de 2000 à 2003 a donné un effet de taille d'ensemble 0.61.

Les enseignants de physique reconnaissent que l'objectif dans la résolution d'un problème de physique n'est pas d'avoir la bonne réponse, mais de comprendre le raisonnement impliqué. De ce fait, la correction tenait compte de quatre critères :

-**Dessin** : Est-ce que l'apprenant a dessiné les vecteurs, axes et corps appropriés? (30%)

-**Définition des variables** : Est-ce que l'apprenant utilise les noms standards des variables ou donne une définition pour les noms non standards? (20%)

-**Équations** : Est-ce que l'apprenant décrit les applications de principes majeurs en écrivant leurs équations sans la substitution des valeurs algébriques et les équations sont-elles correctement utilisées? (40%)

-**Réponse** : Est-ce que l'apprenant a calculé les bons chiffres avec les bonnes unités? (10%)

ANDES2 a été conçu pour augmenter les capacités de compréhension conceptuelles. De ce fait, les apprenants d'ANDES2 devraient avoir de meilleurs résultats par rapport aux critères conceptuels (les 3 premiers). Le tableau ci-après montre l'effet de taille avec la p-valeur de deux t-test consécutifs. Il faut noter que les résultats de 2001 sont non disponibles et qu'il y a eu deux examens d'une heure en 2002, alors leurs résultats sont montrés séparément.

Table 2: Hour exam effect sizes broken down by subscore					
Year	2000	2002a	2002b	2003	Average
Drawings	1.82 (<.001)	0.49 (.003)	0.83 (<.001)	1.72 (<.001)	1.21
Variable definitions	0.88 (<.001)	0.42 (.009)	0.36 (.026)	1.11 (<.001)	0.69
Equations	0.20 (.136)	0.12 (.475)	0.30 (.073)	-0.17 (.350)	0.11
Answers	-0.10 (.461)	-0.09 (.585)	0.06 (.727)	-0.20 (.154)	-0.08

Tableau 5.2.2 Résultats de l'effet des tailles des examens d'une heure par critère (Vanlehn et al., 2005b).

Le résultat qui se dégage de ce tableau est que : au niveau des aptitudes considérées par ANDES2, les apprenants d'ANDES2 avaient de meilleures notes que ceux du contrôle. Ainsi on peut voir que concernant la définition des variables et le dessin, les apprenants ANDES2 ont des notes significativement plus élevées que les apprenants de contrôle chaque année. Ces critères sont des pratiques de résolution qu'ANDES2 oblige à suivre.

En ce qui concerne les équations, il faut savoir qu'avant 2003, ANDES2 était incapable de distinguer un bon d'un mauvais usage des équations. Conséquence, il n'est pas surprenant de constater que les apprenants d'ANDES2 et ceux du contrôle se valent sur ce point avant 2003. En 2003, ANDES2 signalait le mauvais usage des équations avec des points en moins si l'utilisation de leur principe majeur était combinée avec d'autres équations. Toutefois, il n'obligeait pas les apprenants à respecter une méthodologie de résolution. En raison de cela, les performances des apprenants d'ANDES2 sont restées équivalentes à celles du contrôle pour 2003.

Table 6: Residual scores on the 2003 final exam				
	Engineers	Scientists	Others	All
Number of Andes students	55	9	25	89
Number of non-Andes students	178	142	403	823
Andes students mean (stand. dev.)	0.74 (5.51)	1.03 (3.12)	2.91 (6.41)	1.38 (5.65)
Non-Andes students mean (s.d.)	0.00 (5.39)	0.00 (5.79)	0.00 (5.64)	0.00 (5.58)
p(Andes=non-Andes)	0.357	0.621	0.013	0.028
Effect size	0.223	0.177	0.520	0.25

Table 5.2.3 Résultats de l'effet des tailles de l'examen final (VanLehn et Al., 2005b)

Finalement, l'équivalence des notes pour le critère des réponses montre que l'utilisation du résolveur d'équation par les apprenants d'ANDES2 n'a pas altéré leur manipulation algébrique durant l'examen. Par ailleurs, les résultats de l'évaluation de l'examen final (voir le tableau 5.2.3) ont confirmé que les apprenants d'ANDES2 ont de meilleurs résultats que le groupe contrôle.

CHAPITRE VI

ÉVALUATION

Dans le présent chapitre, nous allons mettre en application notre grille d'évaluation avec ANDES2. Pour ce faire, nous allons présenter le contexte dans lequel s'est effectuée l'évaluation d'ANDES2, les résultats de notre évaluation et nous terminerons par une discussion des résultats.

6.1 Contexte de l'évaluation d'ANDES2

Nous avons tenu compte du fait que la démonstration de l'efficacité d'ANDES sur l'apprentissage des apprenants a déjà été effectuée. De plus, les précédentes évaluations se sont acharnées surtout sur la pertinence des stratégies pédagogiques d'ANDES. Notre évaluation d'ANDES sera novatrice dans la mesure où elle permettra de juger des forces et faiblesses de chacune des composantes d'ANDES grâce à l'utilisation de notre grille par un expert.

6.2 Résultats de l'évaluation d'ANDES2

6.2.1 Dimension de l'ergonomique

Toutes les questions concernant cette dimension sont numérotées en commençant par la lettre E pour Ergonomic.

E.1 Guidage

E.1 Guidage

E.1.1 Incitation

Check-list	X	0	1	2	Commentaires
E.1.1.1 Le STI suggère à l'utilisateur des actions permettant de débiter la résolution d'un problème				2	
E.1.1.2 Le STI fournit l'option d'annulation de résolution d'une opération en cours			1		Dans le cas où on demande la résolution d'une équation, il ne nous fait pas confirmer si c'est la bonne équation.
E.1.1.3 Le STI fournit l'option de redémarrer la résolution d'un problème en cours				2	.
E.1.1.4. Le STI fournit de l'aide à l'apprenant				2	
E.1.1.5 Le STI fournit des instructions sonores pertinentes pour guider l'apprenant		0			ANDES2 n'en fait pas usage. Des sons d'encouragement à la fin de la résolution de chaque exercice auraient pu être pertinents.
E.1.1.6 Le STI fournit des instructions vidéo pertinentes pour guider l'apprenant dans la manipulation d'ANDES				2	
E.1.1.7 Le STI fournit des instructions textuelles pertinentes pour guider l'apprenant				2	
E.1.1.8 Le STI fournit un personnage animé attrayant et pertinent qui guide l'apprenant dans la résolution de problème		0			Pas d'utilisation de personnage animé.
E.1.1.9 Il existe une boîte de menu ou un message de feed-back indiquant les choix et le positionnement du curseur				2	
E.1.1.10 L'interface facilite toute sélection, lorsque possible				2	
E.1.1.11 L'interface facilite toute désélection, lorsque possible.				2	

Score total	17/22 =0.77	ANDES2 incite l'apprenant de façon satisfaisante, mais peut s'améliorer avec un plus grand usage des multimédias (personnage virtuel, son, etc...).
-------------	----------------	---

E.1.2 Groupement/Distinction par le format ou le groupement

Check-list	X	0	1	2	Commentaires
E.1.2.1 Le regroupement des items/menus/icônes matche avec le regroupement des tâches				2	
E.1.2.2 L'emplacement des items/icônes/menus est intuitif				2	
E.1.2.3 Les items fréquemment utilisés sont les plus accessibles				2	
E.1.2.4 Chaque item est un membre harmonieux d'une famille d'items				2	
E.1.2.5 Les instructions, les affichages et les messages d'erreurs apparaissent toujours aux mêmes endroits dans le menu				2	
E.1.2.6 Les informations sont présentées par ordre d'importance décroissante				2	
Score	12/12 =1				Les items sont bien organisés dans ANDES2.

E.1.3. Feed-back Immédiat

Check-list	X	0	1	2	Commentaires
E.1.3.1 Il y a une sorte de feed-back pour chacune des opérations de l'apprenant				2	
E.1.3.2 Après la complétion d'une action (ou groupes d'actions) un feed-back indique que la prochaine étape d'actions peut débiter	X				Car cela serait nuisible à l'apprentissage de l'apprenant. Toutefois, ANDES2 fournit ce genre de feed-back uniquement à la demande de l'apprenant.

E.1.3.3 Il y a un feed-back visuel quand un objet est sélectionné ou déplacé				2	
E.1.3.4 Le temps de réponse du système est approprié aux tâches à exécuter				2	
E.1.3.5 Les messages de feed-back sont appropriés et suffisamment précis			1		Pas toujours surtout lorsqu'on fait une erreur dans la dérivation d'une équation, il parle d'erreur d'unité. Il faut alors demander plus d'explication pour comprendre.
Score	7/8 =0.88				ANDES2 fournit un bon feed-back qui peut des fois nécessiter des éclaircissements.

E.1.4. Lisibilité

Check-list	X	0	1	2	Commentaires
E.1.4.1 Les informations sont présentées dans un format facilitant la lecture				2	
E.1.4.2 Le nom de chaque item est explicite des actions qui en découlent				2	
E.1.4.3 Une icône sélectionnée est clairement visible quand elle est entourée d'autres icônes non sélectionnées				2	
E.1.4.4 Le format des items facilite leur lecture				2	
E.1.4.5 Dans un affichage multiple de pages d'écran de données, chaque écran est nommé de façon à décrire sa relation avec les autres			1		La fenêtre d'équations et celle du message d'aide ne sont pas nommées. Cela peut poser une légère gêne si l'on a pas vu la vidéo tutorielle.
E.1.4.6. Les couleurs sélectionnées correspondent aux espérances des codes de couleurs				2	
E.1.4.7 Pour les interfaces de questions/réponses, les questions sont claires et dans un langage simple				2	

E.1.4.8 Les menus sont plus en largeur (plusieurs options dans le menu) qu'en profondeur (plusieurs niveaux d'imbrication de menu)				2	
E.1.4.9 Les instructions en ligne apparaissent en des endroits pertinents sur le(s) écran(s)				2	
Score	17/18 =0.94				ANDES2 utilise un langage clair dans un format lisible et simple à déchiffrer/interpréter

E.2 Charge de travail

E.2.1 Brièveté

Il s'agit ici de la brièveté des actions de manipulation du système.

Check-list	X	0	1	2	Commentaires
E.2.1.1 Les détails excessifs de conception sont évités				2	
E.2.1.2 Les éléments d'entrée et sortie sont suffisamment concis				2	
E.2.1.3 La quantité de fenêtres à gérer en même temps est minimale				2	
E.2.1.4 L'interface utilise entre 4 à 7 couleurs toutes bien éloignées les unes des autres dans le spectre visible				2	
E.2.1.5 Les techniques pour attirer l'attention sont utilisées avec soin sans être excessives				2	
E.2.1.6 L'utilisation des raccourcis clavier et/ou de boutons de raccourcis permet de minimaliser la circulation à travers les menus				2	

E.2.1.7 Le nombre d'icônes est compris entre douze et vingt	0			ANDES2 a plus d'une vingtaine d'icônes, mais le fait qu'il éclaire juste celle dont on a besoin règle ce problème de nombre.
Score	12/14 =0.85			Les actions pour manipuler ANDES2 sont brèves. La seule difficulté est que l'on manipule une vingtaine d'icônes. Toutefois, seules les icônes utiles sont sélectionnables et on n'a pas besoin de toutes les icônes pour résoudre un problème.

E.2.2. Densité informationnelle

Il s'agit ici de réduire la charge informationnelle pour la manipulation de l'interface

Check-list	X	0	1	2	Commentaires
E.2.2.1 Le regroupement des items est intuitif				2	
E.2.2.2 Le nom de chaque item est explicite des actions qui en découlent				2	
E.2.2.3 Tous les items à travers le système sont identifiés par un nom				2	
E.2.2.4 Chaque écran commence avec un titre ou un en-tête décrivant le contenu		0			Les écrans présentant l'énoncé du problème à résoudre, les conseils et les équations ne décrivent pas leur contenu. Il faut avoir vu l'aide vidéo ou consulter l'index d'aide pour comprendre leur rôle.
E.2.2.5 Une légende explicative des couleurs est fournie si les couleurs utilisées sont nombreuses ou difficilement compréhensibles				2	
E.2.2.6 Toutes les actions à travers le système ont des noms significatifs qui ne sont pas ambigus				2	
E.2.2.7 Les informations les plus importantes sont placées en début d'affichage lorsque cela est possible				2	

E.2.2.8 Pour les interfaces questions/réponses, les réponses valides sont listées.				2	
E.2.2.9 Le son est doux pour un feed-back positif et grave dans le cas contraire.	X				ANDES2 n'utilise pas de feed-back sonore. L'utilisation de ce type de feed-back pourrait être irritante dans le cadre d'apprentissage de la physique.
E.2.2.10 Les noms de tous les menus sont judicieux à travers le système				2	
Score	16/18 =0.88				ANDES2 demande très peu d'efforts intellectuels pour la manipulation de son interface. On peut être gêné au début pour comprendre le rôle de chacun des cadres, mais la consultation de la rubrique d'aide permet de contourner cette difficulté.

E.3. Contrôle explicite

E.3.1. Actions Explicites

Le système doit permettre seulement les actions demandées par l'apprenant au moment où ce dernier le désire.

Check-list	X	0	1	2	Commentaires
E.3.1.1 Le temps de réponse est approprié à chaque tâche				2	
E.3.1.2 La vitesse du clavier, mouvement du curseur, sélection de la souris est de 50 à 150 millisecondes				2	
E.3.1.3 Le temps des tâches simples et fréquentes est de 1 seconde				2	
E.3.1.4 Le temps des tâches communes est de 2 à 4 secondes				2	

E.3.1.5 Le temps des tâches complexes est de 8 à 12 secondes				2	
E.3.1.6 Le temps de réponse est approprié au traitement cognitif de l'apprenant				2	
E.3.1.7 S'il y a un temps de traitement d'une durée observable (plus que cinquante secondes), l'apprenant est tenu informé des progrès du système.	X				Durant tous les traitements d'ANDES2, nous n'avons pas noté de temps d'une durée observable.
Score	12/12 =1				ANDES2 répond aux actions de l'apprenant dans les meilleurs délais sans temps d'attente notable.

E.3.2. Contrôle Utilisateur

En tout temps, l'apprenant doit pouvoir contrôler le déroulement des traitements informatiques en cours.

Check-list	X	0	1	2	Commentaires
E.3.2.1 Si le système utilise plusieurs couches de fenêtres, il est facile à l'apprenant de les réarranger sur l'écran	X				Il n'utilise pas plusieurs couches de fenêtres.
E.3.2.2 Si le système utilise plusieurs couches de fenêtres, l'apprenant peut facilement se déplacer d'une fenêtre à l'autre	X				Il n'utilise pas plusieurs couches de fenêtres.
E.3.2.3 Quand une tâche est terminée, le système attend une confirmation de l'apprenant				2	
E.3.2.4 L'apprenant peut annuler des opérations en cours				2	
E.3.2.5 L'apprenant doit confirmer l'exécution de commandes qui ont des conséquences drastiques				2	
E.3.2.6 L'apprenant peut réduire le temps d'entrée des				2	

données en copiant ou en modifiant des données déjà existantes					
E.3.2.7 Si le système a plusieurs niveaux de menus, il y a un mécanisme permettant le retour au menu précédent				2	
E.3.2.8 Si le système permet de reculer au menu précédent, l'apprenant peut modifier les choix précédemment effectués				2	
E.3.2.9 L'apprenant peut avancer et reculer à travers les champs ou boîtes de dialogues d'options				2	
E.3.2.10 Les touches/boutons de fonctions aux conséquences drastiques offrent un moyen d'annulation				2	
E.3.2.11 La continuité de pensée est nécessaire et les informations dont il faut se souvenir à travers les réponses datent de moins de deux secondes				2	
Score	18/18	=1			L'apprenant a le contrôle des opérations en cours et aucune opération drastique n'est effectuée sans sa consultation.

E.4. Adaptabilité

E.4.1. Flexibilité

Check-list	X	0	1	2	Commentaires
E.4.1.1 Les couleurs de l'interface sont modifiables de façon à s'adapter aux goûts et habitudes de l'apprenant		0			L'interface d'ANDES2 est non modifiable.
E.4.1.2 L'interface s'adapte au contexte du problème à résoudre		0			L'interface d'ANDES2 reste la même quelque soit le problème à résoudre. Cette interface est toujours conviviale

				quelque soit le problème à résoudre.
E.4.1.3 Il y a des raccourcis permettant d'éviter l'entrée de longues commandes	X			Les commandes d'ANDES2 sont brèves.
E.4.1.4 Les options de menu sont dépendantes du contexte			2	
E.4.1.5 Un guidage flexible de l'apprenant est fourni		0		
E.4.1.6 Il permet de zoomer l'affichage	X			On n'a pas besoin de zoomer, toutes les images sont simplistes et leur affichage est de dimension lisible.
E.4.1.7 L'apprenant peut nommer les commandes		0		Aucune commande n'est renommable dans ANDES2.
E.4.1.8 L'apprenant peut personnaliser l'affichage des fenêtres			1	L'apprenant peut ajouter ou diminuer le nombre de barres d'outils. Cependant, il ne peut pas déplacer ou modifier l'apparence des cadres utilisés par ANDES2.
E.4.1.9 La séquence de contrôle est flexible			1	L'apprenant peut décider de commencer par dessiner ou définir les variables. Cependant, il doit toujours avoir défini ses variables avant de commencer à écrire ses équations.
E.4.1.10 La conception de la saisie des données est flexible		0		Un espace de trop peut engendrer une erreur de type fatal.
Score	4/16 = 0.25			ANDES2 a une interface rigide laissant un liberté minimale à l'apprenant.

E.4.2. Prise en compte de l'expérience

Check-list	X	0	1	2	Commentaires
E.4.2.1 Le système tient compte de la familiarité, de l'expérience de l'apprenant		0			Novices et expert peuvent effectuer le même nombre d'opérations et ont droit au même nombre de raccourcis.
E.4.2.2 Le niveau de guidance du système diminue avec l'expérience de l'apprenant		0			Le niveau d'aide est le même pour tout niveau d'apprenant.
E.4.2.3 Le niveau de contrôle est adapté au niveau de		0			Le niveau de contrôle ne tient pas compte du niveau de l'apprenant.

l'apprenant (débutant à expert)					
Score	0				ANDES2 ne fait pas de distinction entre un expert et un novice pour la manipulation de son interface. Tous deux ont le même niveau d'accès et de privilèges.

E.5. Gestion des erreurs

Il s'agit ici des erreurs de saisie ou de manipulation, à ne pas confondre aux erreurs survenant lors de la résolution d'un problème.

E.5.1. La protection contre les erreurs

Check-list	X	0	1	2	Commentaires
E.5.1.1 Si la base de données inclut un groupe de données, l'apprenant peut saisir plus qu'un groupe dans le même écran				2	
E.5.1.2 Les points ou les soulignements sont utilisées pour indiquer la longueur des champs		0			Bien que la longueur des champs ne soit pas démesurée, il n'y a pas d'indicateur de longueur de champ dans les formulaires.
E.5.1.3 Le nom du choix de menu de haut niveau est utilisé comme titre dans les menus de bas niveaux			1		Exception faite lorsqu'on accède au sous-menu des formules.
E.5.1.4 Le choix des menus est logique, distinct et mutuellement exclusif				2	
E.5.1.5 Les champs de saisie de données sont vides lorsque possible				2	
E.5.1.6 Lorsque le système affiche plusieurs fenêtres, ces dernières sont visibles et il est simple de naviguer entre elles		0			Impossible. Cela peut être gênant lorsqu'on veut vérifier l'information de la fenêtre précédente pour compléter les informations de la fenêtre courante. Par exemple, lorsqu'on déclare une variable, on n' a plus

				accès à la liste des variables précédemment déclarées.
E.5.1.7 Les boutons/touches de fonctions qui peuvent causer des conséquences sérieuses sont difficilement atteignables			1	Exception du bouton de fermeture du logiciel (x sur la fenêtre) qui est facilement atteignable et peut interrompre la résolution d'un problème en cours et fermer le logiciel.
E.5.1.8 Les boutons/touches de fonctions aux conséquences les plus sérieuses sont éloignés des ceux/celles aux conséquences moins sérieuses ou de ceux/celles fréquemment utilisés			2	
E.5.1.9 Le système prévient l'apprenant des erreurs que ce dernier peut ou va commettre lorsque possible			2	
E.5.1.10 Le système interprète intelligemment la variation de commande de l'apprenant			2	
Score	14/20	=0.70		ANDES2 protège l'apprenant contre les erreurs de saisie. Cependant, il est plus pro-actif que préventif. En d'autres termes, il laisse l'apprenant commettre les erreurs avant de les lui signaler.

E.5.2. La qualité des messages d'erreur

Check-list	X	0	1	2	Commentaires
E.5.2.1 Les messages d'erreurs sont lisibles				2	
E.5.2.2 Concernant la nature de l'erreur commise, les messages d'erreurs sont précis				2	
E.5.2.3 Les messages sont précis en ce qui concerne les actions pour les corriger			1		Pas toujours, car lorsqu'on fait de multiples erreurs il en signale juste une. Or la correction de cette dernière n'annule pas nécessairement les autres.

E.5.2.4 Le système utilise des effets visuels appropriés pour signaler une erreur dans la saisie d'un champ de données				2	
E.5.2.5 Le système utilise des effets sonores appropriés pour signaler une erreur dans la saisie d'un champ de données	X				ANDES2 n'utilise pas d'effets sonores.
E.5.2.6 Les messages d'erreurs évitent l'utilisation de mots violents ou hostiles envers l'apprenant				2	ANDES2 est très courtois dans ses messages d'erreur.
E.5.2.7 Les messages d'erreurs sont formulés de façon à blâmer le système et non l'apprenant		0			Au contraire, bien que ce soit de façon courtoise ANDES2 blâme l'apprenant en cas d'erreur.
E.5.2.8 Le système informe l'apprenant sur la gravité de l'erreur commise		0			ANDES2 ne signale pas les erreurs en terme de gravité.
E.5.2.9 La structure des menus matche avec la structure des tâches				2	Il y a une très forte cohérence entre le menu et les tâches à effectuer.
Score		11/16 =0.69			Les messages d'erreurs sont clairs et courtois, mais tendent à blâmer l'apprenant plutôt que le système et ne tiennent pas compte de la gravité de l'erreur commise.

E.5.3. La correction des erreurs

Check-list	X	0	1	2	<u>Commentaires</u>
E.5.3.1 Le STI permet la correction des erreurs de saisie ou de manipulation				2	
E.5.3.2 Le STI a un système de traçage qui permet à l'apprenant de faire plusieurs retours arrière (« undo ») de saisie ou manipulation				2	
E.5.3.3 Le STI a un système de traçage permettant de faire revenir des saisies ou manipulations annulées (« redo »)				2	
Score		6/6 =1			ANDES2 permet une correction efficace des erreurs de manipulation ou de saisie.

E.6. Homogénéité / Cohérence

Check-list	X	0	1	2	Commentaires
E.6.1 Le format des procédures, des commandes est stable d'un écran à un autre, d'une session à une autre				2	
E.6.2 Les standards industriels ou les normes ont été bien respectés à travers tous les écrans du système	X				Il n'y a pas de standard ou de norme concernant l'interface des STI de physique.
E.6.3 Toutes les icônes sont étiquetées				2	
E.6.4 Il y a des effets visuels saillants permettant d'identifier les fenêtres actives				2	
E.6.5 Toutes les fenêtres ont un titre				2	
E.6.6 S'il y a « quitter » dans le choix de menu, il apparaît toujours en fin de liste				2	
E.6.7 Les instructions en ligne apparaissent à des endroits bien choisis dans tous les écrans du système				2	
Score	12/12 =1				L'interface d'ANDES2 est fortement homogène et cohérente.

E.7. Signifiante des codes et dénomination

Check-list	X	0	1	2	Commentaires
E.7.1.1 La codification et la dénomination respecte celle du domaine enseigné				2	
E.7.1.2 La dénomination et la codification des items sont significatives des actions qui en découlent				2	

Score	4/4 =2	Les codes et dénominations utilisés par ANDES2 respectent le domaine de la physique et sont significatifs des actions qui en découlent.
-------	-----------	---

E.8. Compatibilité

Check-list	X	0	1	2	Commentaires
E.8.1 Le STI présente des similitudes d'interface avec d'autres STI du domaine			1		Oui dans la mesure des actions qu'il permet à l'apprenant de réaliser (définition de variable, équation, vecteur). Toutefois, l'interface en 4 fenêtres pour ressembler à une feuille de papier n'est pas un standard des STI de physique.
E.8.2 Les mots utilisés par le STI sont familiers au domaine				2	
E.8.3 L'organisation des sorties est conforme au champ d'application du système				2	
Score	5/6 = 0.83				Avec ANDES2 les mots et l'organisation des sorties sont conformes à la physique. De plus, il utilise une interface semblable aux STI de physique à la différence qu'il imite la présentation feuille de papier.

6.2.2 Dimension utilité

U.1.Le modèle apprenant: fiabilité

Check-list	X	0	1	2	Commentaires
U.1.1 Le modèle apprenant est explicitement défini dans le système		0			Il n'y a pas de modèle apprenant explicitement défini dans le système. Toutefois, le système évalue la capacité

				de résolution d'un problème par l'apprenant.
U.1.2 Le modèle apprenant est visible par l'apprenant lui-même	X			
U.1.3 Le modèle apprenant peut se calibrer automatiquement	X			
U.1.4 L'apprenant peut calibrer personnellement le modèle apprenant	X			
U.1.5 Le calibrage permet de se faire une idée de(s) apprenants(s) dès le départ	X			
U.1.6 L'information diagnostiquée obtenue par le système sur les connaissances de l'apprenant est identique aux diagnostics obtenus indépendamment du système à tout moment			1	Oui, sauf dans le cas où l'apprenant fait deux erreurs dans une même équation. ANDES2 va identifier une erreur à la fois.
U.1.7 Le système tient compte des erreurs récurrentes de/ des apprenants (modèle cognitif : overlay par perturbation)		0		Les erreurs récurrentes de l'apprenant ne sont pas prises en compte lors de l'analyse de son raisonnement.
U.1.8 Le système tient compte des informations affectives pouvant influencer le comportement de/des apprenants (modèle affectif de l'apprenant)		0		Les émotions de l'apprenant ne sont pas prises en compte lors de l'analyse de son raisonnement.
U.1.9 Le système tient compte du contexte d'apprentissage			2	
U.1.10 Le système fait un traçage du raisonnement de/des apprenants (modèle tracking)			2	
U.1.11 Le système analyse les événements d'apprentissage pour déterminer les connaissances qui y sont employées (knowledge tracking)			2	
U.1.12 Les apprenants peuvent collaborer entre eux pour résoudre un problème (approche collaborative)	X			Dans le cadre du cours où est utilisé ANDES2, il n'y a pas de travaux de groupe. Par conséquent, la collaboration n'est pas une option obligatoire.

Score	7/14 =0.50	Bien qu'ANDES2 ne définit pas explicitement un modèle apprenant, il est capable d'inférer sur le raisonnement de ce dernier. Toutefois, il est limité par le fait qu'il ne tient compte ni des erreurs récurrentes ni des émotions de l'apprenant.
-------	-------------------	--

U.2.Le modèle pédagogique

U.2.1 Précision/présentation des objectifs pédagogiques et identification de l'approche pédagogique utilisée:

Check-list	X	0	1	2	Commentaires
U.2.1.1 Le système présente de façon précise ces objectifs pédagogiques				2	
U.2.1.2 Le système intervient lorsque l'apprenant s'éloigne de la solution (coaching)				2	
U.2.1.3 Le système interroge l'apprenant sur ses erreurs afin de le faire progresser (approche socratique)	X				Une approche socratique aurait une autre façon pertinente d'implanter l'aide dans ANDES2.
U.2.1.4 Le système essaye d'induire l'apprenant en erreur (approche par perturbation)	X				Non, c'est plutôt le contraire.
U.2.1.5 Le système permet à l'apprenant d'apprendre en modifiant des paramètres et en observant les conséquences de ses actions dans un environnement simulé (guidage)	X				ANDES2 ne fait pas de simulation.
U.2.1.6 Le ou les approches pédagogiques utilisées par le système sont pertinentes par rapport au domaine enseigné				2	Oui car le "next step help" permet à l'apprenant de savoir quelle est la prochaine étape à suivre tandis que le "what's wrong help" l'aide à comprendre ses erreurs.
U.2.1.7 Le tuteur peut changer d'humeur en fonction de la qualité du raisonnement de	X				ANDES2 n'utilise pas de tuteur virtuel.

l'apprenant (réalité virtuelle)					
U.2.1.8 Si le système a plusieurs approches pédagogiques, l'alternance des approches est pertinente	X				ANDES2 ne possède pas plusieurs approches pédagogiques pour en faire l'alternance de façon pertinente ou non.
Score	6/6 =1				ANDES2 présente ses objectifs pédagogiques à l'aide d'une vidéo de présentation du système. Il utilise une approche pédagogique de style coaching qui est adéquate.

U.2.2 Adéquation des scénarios didactiques aux objectifs pédagogiques et exactitude du diagnostic:

Check-list	X	0	1	2	<u>Commentaires</u>
U.2.2.1 Les scénarios didactiques sont conformes aux objectifs pédagogiques du système				2	
U.2.2.2 Les scénarios didactiques tiennent compte de l'évolution des connaissances/habiletés de l'apprenant		0			Non car qu'importe son niveau, l'apprenant peut décider de résoudre des problèmes plus complexes ou non sans que le système n'intervienne.
U.2.2.3 Les scénarios didactiques tiennent compte de la régression des connaissances/habiletés de l'apprenant		0			Non car qu'importe son niveau, l'apprenant peut décider de résoudre des problèmes moins complexes ou non.
U.2.2.4 Les scénarios didactiques tiennent compte du contexte d'apprentissage				2	
U.2.2.5 Les diagnostics du modèle pédagogique sont réalistes				2	
Score	6/10 =0.6				Les scénarios didactiques sont conformes aux objectifs pédagogiques d'ANDES2. Étant donné qu'il a été prévu qu'il soit utilisé dans le cadre d'un cours avec un enseignant, les scénarios didactiques d'ANDES2 ne tiennent pas compte de l'évolution ou de la régression des connaissances de l'apprenant.

U.3. La Base de connaissance : conformité au domaine

Check-list	X	0	1	2	Commentaires
U.3.1 Le système utilise des connaissances procédurales (si ... alors ... sinon)				2	
U.3.2 Le système utilise des connaissances déclaratives (faits, concept, évènement, images, proposition, ...)				2	
U.3.3 Le système utilise des connaissances épisodiques	X				Les connaissances épisodiques étant un nouveau concept, on n'a pas encore la preuve de leur pertinence par rapport à l'enseignement de la physique.
U.3.4 Les connaissances sont retranscrites à l'aide d'un formalisme logique (règle de production, logique floue, ...)				2	ANDES2 est constitué de 356 problèmes et de 550 règles de physique.
U.3.5 Les connaissances sont retranscrites à l'aide d'un formalisme sémantique (réseaux sémantiques)				2	À chaque problème d'ANDES2 est associé un graphe de solutions.
U.3.6 Les connaissances sont retranscrites à l'aide d'un formalisme hybride (réseaux bayésiens)	X				ANDES2 n'utilise pas les réseaux bayésiens parce qu'il est utilisé dans le cadre d'un cours et c'est l'enseignant qui décidera de la progression de l'apprenant.
U.3.7 Les types de connaissances et les formalismes utilisés sont en accord avec les scénarios didactiques				2	
U.3.8 Le corpus du modèle de connaissance peut évoluer				2	Les enseignants peuvent ajouter des problèmes dans la base de connaissance d'ANDES2.
U.3.9 On a accès aux connaissances du système				2	Car le raisonnement de l'apprenant est comparé au système expert.
U.3.10 Le système peut partager ses connaissances avec d'autres systèmes			1		Oui mais seulement avec ceux qui utilisent le même système expert que lui.
Score	15/16 =0.94				ANDES2 utilise une base de connaissance adéquate pour la satisfaction des objectifs pédagogiques. Toutefois, il ne peut partager ces connaissances qu'avec des STI ayant le

		même système expert que lui.
--	--	------------------------------

U.4. La modèle expert : Rectitude

Check-list	X	0	1	2	<u>Commentaires</u>
U.4.1 Le système propose des solutions exactes aux problèmes proposés				2	
U.4.2 Le système détermine de façon précise le(s) cause(s) de(s) erreur(s) de l'apprenant			1		Il peut arriver que le système ne soit pas capable de déterminer la source de l'erreur. Dans ce cas, il demande des précisions à l'apprenant.
U.4.3 Le système est capable de prédire les erreurs de l'apprenant à partir d'un début de raisonnement de l'apprenant		0			ANDES2 attend que l'erreur soit commise avant de la signaler à l'apprenant.
U.4.4 On a accès aux connaissances du système expert (l'expert n'est pas une boîte noire)				2	
U.4.5 Les connaissances de l'apprenant sont comparées à celle du système expert (boîte de verre)				2	
U.4.6 Ce modèle est basé sur des théories de cognition pertinentes	X				ANDES2 n'utilise pas de théorie de cognition car il se base sur l'utilisation d'un modèle expert.
Score	7/10 =0.70				ANDES2 utilise un modèle expert qui sait résoudre les problèmes du domaine, mais est limité par le fait qu'il n'anticipe pas les erreurs probables.

6.3 Discussion de l'évaluation

Pour discuter du résultat de notre évaluation, nous allons analyser le pourcentage de satisfaction obtenue par ANDES2 pour chacun des critères de la dimension ergonomique et

utilitaire de notre grille. Nous terminerons par une étude comparative de notre évaluation à celle faite empiriquement par l'école navale pendant cinq ans.

6.3.1 Dimension ergonomique

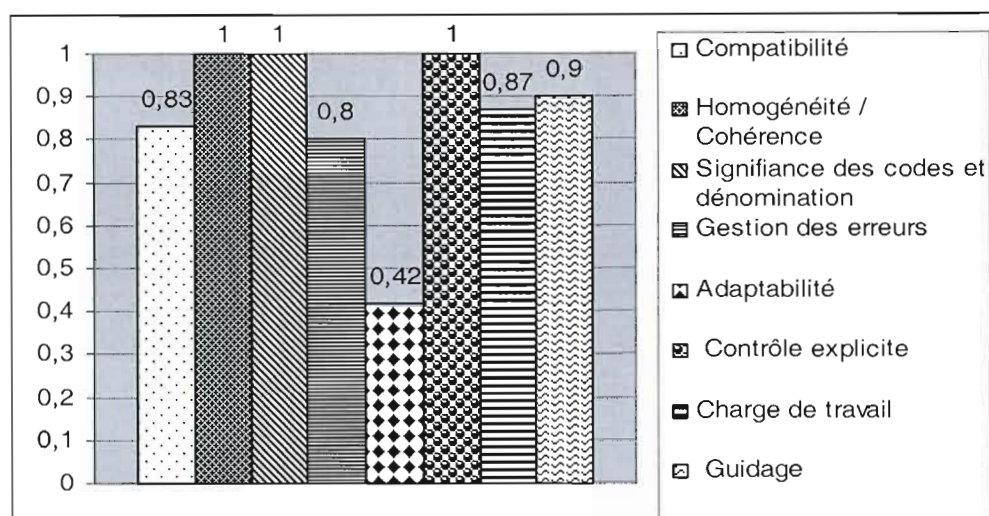


Figure 6.3.1.1 histogramme récapitulatif de l'évaluation de l'utilité d'ANDES 2

L'évaluation de cette dimension nous a permis de constater qu'ANDES2 est très complet sur le plan ergonomique (voir figure 6.3.1.2). Bien que n'utilisant pas un grand éventail multimédia, ANDES2 fournit tout le guidage nécessaire à sa bonne manipulation. Il exige une charge de travail minimale pour sa manipulation et offre un bon niveau de contrôle à l'apprenant. Par ailleurs, ANDES2 offre une présentation homogène des informations dont les dénominations et les signifiants sont conformes à la matière enseignée. ANDES2, bien qu'innovateur dans la présentation de son interface (ressembler à une feuille d'exercice de physique), présente des similitudes avec les autres interfaces du domaine (rédaction d'équations, utilisations des vecteurs, etc...). Enfin, il offre un service minimum permettant d'éviter les erreurs de manipulations.

Toutefois, fort est de constater qu'ANDES2 présente des lacunes ergonomiques qu'on ne saurait négliger. En effet, il présente une interface rigide que l'utilisateur ne peut personnaliser. Il ne tient pas compte de l'expérience de l'apprenant dans la manipulation de

l'interface à l'exception du fait qu'on n'a pas besoin d'écouter la vidéo d'introduction si on l'a déjà fait.

6.3.2 Dimension de l'utilité

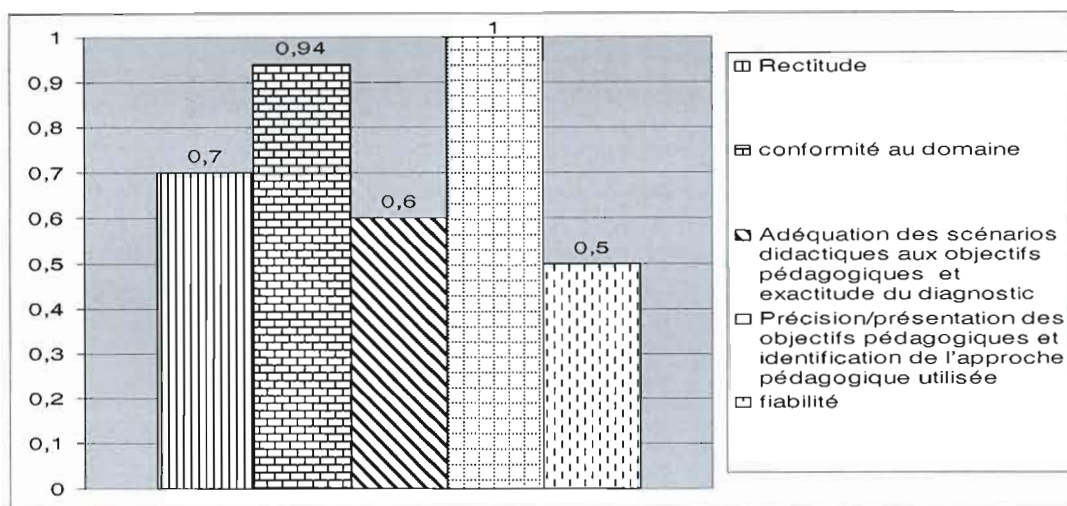


Figure 6.3.2.1 histogramme récapitulatif de l'évaluation de l'utilité d'ANDES 2

Sur le plan de l'utilité, notre évaluation nous a permis de constater qu'en général, les composantes d'ANDES2 satisfont aux critères qu'on attend de chacune d'elles (voir figure 6.3.2.1). Cela confirme le résultat obtenu par les évaluations empiriques d'ANDES2 (voir section 5.2). Ainsi en ce qui concerne le modèle expert nous avons pu constater qu'ANDES2, bien que n'utilisant pas des théories de la cognition, et en se servant d'un système expert, est capable de résoudre les problèmes posés et de détecter les erreurs de l'apprenant lorsqu'elles surviennent. Le modèle expert d'ANDES2 raisonne en s'appuyant sur une base de connaissance qui n'inclut pas les erreurs récurrentes de l'apprenant. Cette base de connaissance utilise un type de connaissances conformes au domaine et qui sont représentées sous forme de graphes de solutions et des règles de production. De plus, cette base à contenu dynamique peut être partagée de façon limitée. Les enseignants peuvent ajouter de nouveaux problèmes à résoudre à la base de connaissance d'ANDES2. La base de connaissance d'ANDES2 peut être partagée avec les STI utilisant le même système expert. Le modèle

pédagogique d'ANDES2 qui interagit avec cette base de données est lui-même d'assez bonne qualité.

En effet, ce modèle pédagogique présente adéquatement les objectifs pédagogiques à l'aide d'une vidéo d'introduction. Ce modèle utilise une approche pédagogique de style coaching. Cela s'explique par le fait qu'il n'interrompt pas l'apprenant dans sa résolution de problèmes, mais lui signale les erreurs de raisonnement en colorant les entrées en rouge si les entrées sont erronées et en vert sinon. De plus, il fournit plusieurs niveaux d'aide permettant la résolution des problèmes proposés (« next step help » et « what's wrong help »). Toutefois, le fait qu'ANDES2 est supposé être utilisé dans le cadre d'un cours avec un enseignant (et non de façon autodidacte), ses scénarios didactiques ne tiennent pas compte de l'évolution ou de la régression des connaissances de l'apprenant. Ainsi, c'est à l'apprenant de choisir le problème qu'il souhaite résoudre indépendamment de son niveau de connaissances.

Le modèle apprenant est le modèle le plus négligé d'ANDES2. En effet, il n'existe pas de modèle apprenant à proprement parler dans ANDES2. Malgré cela, il est capable d'inférer sur le raisonnement de l'apprenant en comparant le raisonnement de l'apprenant à travers les graphes de résolution de problème et de son système expert. Cependant, il ne tient pas compte des erreurs récurrentes ou des émotions pouvant influencer le raisonnement de l'apprenant. La fiabilité de ce modèle apprenant s'en trouve fragilisée. Par ailleurs, force est de constater qu'ANDES2 ne permet pas un travail collaboratif entre les apprenants.

En définitive, nous pouvons dire que l'évaluation d'ANDES2 nous aura permis de faire ressortir les qualités et les lacunes de chacune des composantes.

6.3.3 Analyse comparatives des résultats de notre évaluation d'ANDES2

Dans cette section, nous allons comparer les résultats de notre évaluation par rapport aux résultats obtenus par l'école navale (voir section 5.2).

Les évaluations empiriques d'ANDES2 ont principalement porté sur sa capacité à améliorer les aptitudes des apprenants (méthodologie et capacité de résolution) et très peu sur

l'aisance de ces derniers à manipuler ANDES2. Notre évaluation d'ANDES2 sur le plan ergonomique aura permis de combler ce manque. Cela aura pour effet de faire ressortir le fait qu'ANDES2 est très ergonomique, mais souffre d'une interface rigide et tient très peu compte de l'expérience de l'apprenant.

Par ailleurs, il est apparu des évaluations empiriques d'ANDES2 qu'il améliore la méthodologie et la capacité de résolution des problèmes de physiques de ces apprenants. Ces résultats convergent avec notre évaluation sur le plan de l'utilité d'ANDES2. En effet, notre évaluation aura permis de montrer que les composantes d'ANDES2, exception faite du module de l'apprenant, satisfont aux critères de l'utilité (critères qui déterminent la capacité du STI à améliorer les aptitudes des apprenants). De plus, elle aura permis de montrer qu'ANDES2 n'encourage pas la collaboration entre les apprenants.

D'une manière générale, les évaluations empiriques d'ANDES2 et la nôtre, au moyen de notre grille d'évaluation, sont convergentes et complémentaires. Plus précisément, notre évaluation aura permis d'identifier les forces des composantes d'ANDES2 qui expliquent le résultat des évaluations empiriques obtenus.

CHAPITRE VII

CONCLUSION

7.1 Synthèse

Les systèmes tutoriels intelligents sont des systèmes ayant pour objectif de permettre à un apprenant ou un groupe d'apprenants d'acquérir de nouvelles connaissances en s'adaptant à leur comportement. De par leur objectif, on s'attendrait à ce que ces systèmes soient très répandus dans le marché scolaire, ce qui n'est pas le cas. De plus, comme l'a remarqué Sharon Answorth en 2005, les systèmes tutoriels intelligents sont « mourants ». Le problème est de savoir quels sont les critères auxquels doit répondre un STI pour pouvoir augmenter son intégration au milieu scolaire.

Afin de résoudre ce problème, nous avons proposé une approche novatrice à savoir : une grille d'évaluation bidimensionnelle par inspection d'un STI. Les deux dimensions attaquées par la grille sont l'utilisabilité (l'ergonomie) et l'utilité d'un STI. Ces deux dimensions ont été choisies dans la mesure où elles permettent d'avoir une vision complète de la bonne qualité du STI. D'une part, l'ergonomie permet de juger de l'aisance de manipulation du système. Cette dimension s'intéresse particulièrement à l'interface de communication. D'autre part, l'utilité du système permet de juger de la pertinence de la stratégie pédagogique du système. Ce dernier critère concerne les quatre autres composantes classiques d'un STI à savoir : modèle apprenant, modèle pédagogique, base de connaissance et modèle expert. Pour élaborer notre grille nous nous sommes basés sur les critères de Bastien et Scapins tels que suggérés par la littérature pour la dimension ergonomique. Pour la dimension utilitaire, nous sommes servis d'une approche modulaire telle que suggérée par Marc et Greer (1993).

Les critères mentionnés par Trico et Al (2003) ont permis l'élaboration du questionnaire pour chacune des composantes à évaluer sur le plan utilitaire.

Par la suite, afin de tester la pertinence de notre grille d'évaluation, nous l'avons appliqué à un STI mature bien connu et accessible à savoir ANDES version 2. Les précédentes évaluations de ce dernier ont montré qu'il était dans son ensemble un atout indéniable facilitant l'apprentissage. Notre évaluation quant à elle en plus de confirmer sa maturité a été révélatrice dans la mesure où elle a su faire ressortir de façons plus précises les atouts et les faiblesses d'ANDES2. Sur le plan ergonomique, il est ressorti qu'ANDES2 bien que suffisamment intuitif, manquait d'adaptabilité. En effet, c'est un système très peu personnalisable qui ne tient pas en grande considération le niveau de l'expérience de l'utilisateur du système qu'est l'apprenant. Sur le plan utilitaire, il est ressorti que la composante de moins bonne qualité d'ANDES2 est le modèle apprenant. En effet, cette composante n'a pas été explicitement développée dans ANDES2. Ainsi, bien qu'ANDES2 soit capable d'inférer sur les connaissances de l'apprenant à partir de la comparaison du raisonnement de l'apprenant à un graphe de résolution, il ne bénéficie pas du profil de l'apprenant permettant d'assurer une meilleure validité du diagnostic du modèle pédagogique.

7.2 Perspectives et limites de notre grille d'évaluation

Notre grille d'évaluation est limitée par le fait qu'elle ne peut être utilisée que par un expert en ce qui concerne l'évaluation de l'utilité. Aussi, notre grille ne peut permettre d'évaluer l'efficacité de l'apprentissage de(s) apprenant(s) qui ne peut être évalué qu'empiriquement. Toutefois, cette limite peut être compensée par la manière dont la grille d'évaluation sera utilisée.

En effet, la juxtaposition antérieure ou postérieure d'une évaluation empirique avec une évaluation à l'aide de notre grille permet de combler ce manque. C'est d'ailleurs l'une des

raisons pour lesquelles nous avons choisi d'appliquer notre grille à ANDES2 qui est un STI ayant déjà subi cinq ans d'évaluation empirique.

Par ailleurs, l'évaluation de STI à l'aide de notre grille est intéressante sur un double aspect. D'une part, par le fait qu'elle fait ressortir les points forts d'un STI, elle pourra servir à leur classification en termes d'ergonomie et d'utilité. L'évaluation d'ANDES2 a été faite dans cette perspective. D'autre part, si le STI est en cours de réalisation, elle peut servir à déterminer de façon précise où l'emphase doit être mise en termes d'effort de développement.

Enfin, il est important de noter que la liste de critères pour satisfaire au plan ergonomique et utilitaire de notre grille d'évaluation est complète, mais non exhaustive. D'autres critères peuvent lui être rajoutés. Ainsi, par exemple, le critère d'accessibilité peut lui être rajouté sur le plan ergonomique. Cela s'explique par le fait que les systèmes tutoriels intelligents sont soumis à une démocratisation des systèmes d'information comme exigés de plus en plus par plusieurs normes (par exemple les systèmes gouvernementaux).

BIBLIOGRAPHIE

- Ainsworth, Shaaron. 2005. «Evaluation Methods of Learning Environments». En ligne <http://www.psychology.nottingham.ac.uk/staff/Shaaron.Ainsworth/aied_tutorialsides2005.pdf>. Dernière modification October 5, 2006. Consulté le 25 Août 2007.
- Aleven, V., O. Popescu et K. Koedinger. 2001. «Pedagogical content knowledge in a tutorial dialogue system to support self-explanation». In *Papers of the AIED-2001 Workshop on Tutorial Dialogue Systems*, p. 59-70
- Aleven, V., O. Popescu et K. Koedinger. 2003. «A Formative Classroom Evaluation of a Tutorial Dialogue System that Supports Self-Explanation. ». In *Supplemental Proceedings of the 11th International Conference on Artificial Intelligence in Education, AIED*, volume VI
- Amiel, A., J-F Camps, G. Lutz, F. PLEGAT-Soutjis et A.m. Tricot. 2002. Acceptabilité de form@lion : Évaluation et recommandation. Rapport d'étude CERFI. IUFM de Midi-Pyréné: 25 p.
- Bastien, D J. M. C., et D.L. Scapin. 1993. «*Ergonomic criteria for the evaluation of human-computer interfaces*». Rocquencourt, France, INRIA. Vol 156
- Bastien, D J. M. C., et D.L. Scapin. 1997. «Ergonomic criteria for evaluating the ergonomic quality of interactive systems». *BEHAVIOUR & INFORMATION TECHNOLOGY*. vol. 16, no NO. 4/5, p. 220-231.
- Bastien, J.M.C, D.L Scapin et C. Leulier. 1999. «The ergonomic criteria and ISO/DIS 9241-10: a pilot comparaison in an evaluation task». *Interacting with computers*. vol. 11 no 3, p. 299-322.
- Beck, Joseph, M. Stern et E. Haugsjaa .2004. Applications of AI in Education. Crossroads the ACM student Magazine. Xrds3-1. En ligne <<http://www1.acm.org/crossroads/xrds3-1/aied.html#R32>>. Consulté le 25 Août 2007.
- Brown, J S., R. R. Burton et A. G Bell. 1975. «A step toward creating a reactive learning environment». *International Journal of Man-Machine Studies*. vol. 7, p. 675.
- Carbonell, R.J. 1970. «AI in CAI: An artificial intelligence approach to computer-assisted instruction». *IEEE Transactions on Man-Machine Systems*. vol. 11, p. 190-202.

- Cheung, B, L. Hui, J. Zhang et S.M. Yiu. 2003. «Smart tutor: an intelligent tutoring system web-based adult education». *The journal of systems and software*. vol. 68, p. 11-25.
- Clancey, W.J. 1987. Knowledge-Based Tutoring : The GUIDON Program. Cambridge Mass. : MIT Press.
- Dubois, Daniel. 2004. «Reproduction de mécanismes de "conscience" dans un agent logiciel : Application à un agent d'enseignement intelligent ».Thèse de doctorat, Montréal, Université du Québec à Montréal, 10- 38 p.
- Frebourg, Laure. 2001. « La représentation des connaissances en Intelligence Artificielle ». DEA Méthodes Scientifiques de Gestion, Paris, Ecole Polytechnique Ecole Nationale Supérieure des Mines de Paris, Université Paris IX-Dauphine, 18 p.
- Gagne, R. M. , Briggs L. J. et Wager W. W 1988. Principles of Instructional Design, 3^e ed. Chicago: Holt, Rinehart and Winston, 232 p.
- Gouadjo-AKO, P. F. 2005. « un outil d'assistance pour le design pédagogique : conception et implémentation du CIA ». Mémoire de Maîtrise, Montréal, Université du Québec à Montréal, 1,12 p.
- Gupta, A., et P. S. Grover. 2004. « Proposed Evaluation Framework for Adaptive Hypermedia». In *Proceedings of Third Workshop on Empirical Evaluation of Adaptive Systems,AH2004*, Eindhoven University of Technology, The Netherlands.
- Hume, G., J. Michael, A. Rovick et M. Evens. 1996. « Hinting as a tactic in one-on-one tutoring». *Journal of the Learning Sciences*. vol.,5, no 1, p. 23-49.
- Iqbal A., R. Oppermann, A. Patel et A. Kinshuk. 1999. «Classification of Evaluation Methods for Intelligent Tutoring Systems, Software Ergonomie ». eds: *Arend, U. et al., Teubner Stuttgart*, p. 169-181.
- Jean, S. 2000. «Application de recommandations ergonomiques». *Actes des Rencontres Jeunes Chercheurs en IHM*, p. 39-42.
- Klopfer, L.E. 1986. «Intelligent tutoring systems in science education: The coming generation of computer-based instructional programs». *Journal of Computers in Mathematics and Science Teaching*. vol. 5, p. 16-32.
- Kulik, C., Kulik, J. et R. Bangert-Drowns. 1990. « Effectiveness of mastery learning programs: A metaanalysis.». *Review of Educational Research*. vol. 60, no 2, p. 265-306.
- Legree, P. J., Gillis P. D. et Orey M. A. 1993. «The quantitative evaluation of intelligent tutoring system application: Product and process criteria». *Journal of Artificial Intelligence and Education*. vol. 4, no 2/3, p. 209-226.

- Lesgold, A., S. Lajoie, M. Bunzo et G. Eggan. 1992. «Sherlock: A coached practice environment for an electronics troubleshooting job. In J. Larkin et R. Chabay ». In *J. Larkin et R. Chabay (Eds.), Computer Based Learning and Intelligent Tutoring*. Hillsdale, NJ: LEA, p. 202-274.
- Lewis Johnson, W., et Beal C. 2005. «Iterative Evaluation of a Large-Scale, Intelligent Game for Language Learning Artificial intelligence in education ». p. 290-297.
- Mark, M. A., et Greer. 1993. «Evaluation methodologies for intelligent tutoring systems». *Journal of Artificial Intelligence and Education*, p. 129-153.
- Meisel, H., et E. Compatangelo. 2004. «An ontology-based architecture for the design of knowledge bases in Intelligent Instructional Systems». *Int'l Jour. of Interactive Technology and Smart Education*. vol. 1, no 3, p. 5-19.
- Merrill, D. C., B. J. Reiser, M. Ranney et J. G. Trafton. 1992 «Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems». *The Journal of the Learning Sciences*. vol. 2, no 3, p. 277-306.
- Murray, T. 1993. «Formative qualitative evaluation for `exploratory' ITS». *Research.Journal of Artificial ntelligence and Education*. p. 179-207.
- Nedic, Z., V. Nedic et J. Machotka. 2002. «Intelligent Tutoring System for teaching 1st year engineering». *UICEE World Transactions on Engineering and Technology Education*. vol. 1, no 2, p. 241-243.
- Nielsen, J. 1993. «usability Engineering ». Boston, MA: AP Professional. Hardcover: ISBN 0-12-518405-0 Softcover: ISBN 0- 12-518406-9
- Nkambou, R. 1997. «Using Fuzzy logic in ITS-course generation». In *9th International Conference on Tools with Artificial Intelligence* (November 3-8): IEEE Computer Society.
- Nkambou, R. 1998. «Knowledge engineering with fuzzy logic in an industrial intelligent training system». In *the 3rd Annual International Conference on Industrial Engineering Theories, Applications and Practice* (December).
- Nkambou, R. 2005. « notesde cours inf7470, scéance 3 ».En ligne
<<http://gdac.dinfo.uqam.ca/inf7470/>> Consulté le 25 Août 2007.
- Nogry, S., S. Jean-Daubias et M. Ollagagnier-Beldame. 2004. «Évaluation des EIAH : une nécessaire diversité des méthodes ». Dans *TICE* (Novembre) p. 265-271.

- O'Neil, H., et E. Baker. 1987. «Issues in intelligent computer aided instruction: Evaluation and measurement». In *The Computer as Adjunct to the Decision Making Process*: Conoley J. C.
- Papert, S. 2000. Logo foundation .En ligne.
<<http://el.media.mit.edu/Logo-foundation/logo/index.html>>. Consulté 25 Août 2007.
- Paris, S.G., M.Y. Lipson et K.K. Wixson. 1983. «Becoming a strategic reader.». *Contemporary Educational Psychology*. vol. 8, p. 293-316.
- Parry, J. D., et A. M. Hofmeister. 1986. «The development and validation of an expert system for special educators». *Learning Disability Quarterly*, p. 124-132.
- Partridge, D.. 1986. Artificial Intelligence: Applications in the future of software engineering. Ellis Horwood Series in Artificial Intelligence, J. Campbell, West Sussex, England: ed., Ellis-Horwood
- Pierotti, D. 1996. How to conduct a Heuristic Evaluation A System Check-list. Usability Analysis and Design. Xerox Corporation
<<http://www.stcsig.org/usability/topics/articles/hechecklist.html> >. Consulté 25 Août 2007
- Quillan, R. 1968. « Semantic memory. Semantic Information Processing ». Cambridge Mass. : MIT Press., p. 216-270.
- Retschitzki, J. 1995. Apprendre avec l'informatique: vers une révolution ? , Institut de psychologie, Université de Fribourg (Suisse) En ligne.
<<http://www.unifr.ch/spc/UF/95juin/retschitzki.html>>
- Shapiro, J. A.. 2005. «Algebra subsystem for an intelligent tutoring system.». *International Journal of Artificial Intelligence in Education*. Vol. 15, no 3, p. 205-228.
- Shute, V.J.. 1993. « A macroadaptive approach to tutoring ». *Journal of Artificial Intelligence in Education* vol. 4, no 1, p. 61-93.
- Sleeman, D., A. E. Kelly, Martinak R., R. D. Ward et J. L. Moore. 1989. «Studies in the diagnosis and remediation of high school algebra students. ». *Cognitive Science*. vol. 13, p. 551-568.
- Tchetagni, J., R. Nkambou et J. Bourdeau. 2005. «Supporting student reflection in an intelligent tutoring system for logic programming». J. Kay, A. Lum & J.-D. Zapata-Rivera (eds.), *Artificial Intelligent in Education, 12th International Conference, AIED2005, Proceedings of the Workshop on Lomore*, p. 42-51.

- Tchounikine, P. , M. Baker, N. Balacheff, M. Baron, A. Derycke, D. Guin, J-F. Nicaud et P. Rabardel (2004). Platon-1 : quelques dimensions pour l'analyse des travaux de recherche en conception d'EIAH. Rapport d'Action Spécifique du CNRS, 1-19p.
- Tchounikine, P., M. Baker, N. Balacheff, M. Baron, A. Derycke, D. Guin, J-F. Nicaud et P. Rabardel. 2002. «Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain». *Revue I3*. vol. 2 no 1, p. 36.
- Tricot, A., F. Plégat-Soutjis, J-F. Camps, A. Amiel, G. Lutz et A. Morcillo. 2003. «Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH ». *EIAH*, p. 391-402
- Vanlehn, K. 1990. «Mind Bugs: The Origins of Procedural Misconceptions. ». *Cambridge, MA: MIT Press*.
- Vanlehn, K., C. Lynch, K. Schulze, J.A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein et M. Wintersgill. 2005a. « The Andes Physics Tutoring System: Lessons Learned». *International Journal of Artificial Intelligence and Education*. vol. 15 no (3).
- Vanlehn, K., C. Lynch, K. Schulze, J. A. Shapiro, R. H. Shelby, L. Taylor, D. J. Treacy, A. Weinstein et M. C. and Wintersgill. 2005b. The Andes physics tutoring system: Five years of evaluations. *Proceedings of the Artificial Intelligence in Education Conference*. (Amsterdam). In: G. I. McCalla and C.-K. Looi IOS